Estimation of Max-Stable Processes Using Monte Carlo Methods with Applications to Financial Risk Assessment

Francisco Chamú Morales

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics.

Chapel Hill
2005

Approved by

Advisor: Richard L. Smith

Reader: Yuguo Chen

Reader: Harry L. Hurd

Reader: Eric Ghysels

Reader: Chuanshu Ji

ABSTRACT

FRANCISCO CHAMÚ MORALES: Estimation of Max-Stable Processes Using Monte Carlo Methods with Applications to Financial Risk Assessment
(Under the direction of Richard L. Smith)

Multivariate extreme value theory is concerned with the joint distribution of extremes of multiple random variables. The theory is used in a number of areas such as finance and environmental science. For example, empirical observations suggest that extreme events in financial time series occur in clusters and are dependent across different assets.

It is possible to characterize the extremal behavior of a multivariate stationary time series in terms of a limiting max-stable process. Our approach for the statistical modeling of max-stable processes is based on Moving Maxima (MM) processes, and a multivariate extension known as Multivariate Maxima of Moving Maxima (M4) processes.

This work is concerned with developing Monte Carlo methods for filtering, prediction, and parameter estimation of M4 processes. The model is a state-space representation, where the state is an unobserved M4 process, and the observed process is a nonlinear transformation of the state with additive Gaussian noise.

Our contributions can be divided in three areas. First, we show that two special cases of moving maxima processes, which we refer to as MM(1) and MM(2) processes, are second- and third-order Markov, respectively. Second, we propose sequential Monte Carlo methods to approximate the filtering distributions of M4 processes. Third, we provide Markov chain Monte Carlo methods for obtaining the posterior distribution of the unknown parameters in the model.

Accurate financial risk assessment is of major interest for financial institutions and regulators. For instance, the Basel committee recommends reporting Value at Risk for a 10-day holding period. We propose simulation-based prediction of M4 processes for estimating market risk over a multiple-day period, and apply this method to four Nasdaq sector indices.

ACKNOWLEDGMENTS

I would especially like to thank my advisor, Richard L. Smith, for his generous advice and support, for many illuminating discussions, and for his sharp comments on the manuscript. It has been a privilege working with him.

I wish to express my sincere gratitude to the members of my committee: Yuguo Chen, Harry L. Hurd, Eric Ghysels, and Chuanshu Ji, for their suggestions that led me to improve this work. I have truly enjoyed interacting with each one of them.

I have been fortunate to work as research assistant for Robert N. Rodriguez. I greatly appreciate his support and understanding, which allowed me to fully concentrate on the dissertation when it was needed.

I am most grateful to my wife, Jeniffer Iriondo Pérez, for her endless support and encouragement during all the stages of this project, for her patience when we had to spend time apart, and for carefully proofreading the manuscript.

I am deeply indebted to my family, who has always supported me, and especially to my parents, for their love, guidance, and vision.

Thanks to the faculty and staff of our Department, and to my fellow graduate students, for a great experience in Chapel Hill.

I gratefully acknowledge financial support from the Fulbright-García Robles program and Consejo Nacional de Ciencia y Tecnología (CONACYT).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

A'	transpose of A
AR	autoregressive
ARCH	autoregressive conditional heteroskedastic
Cov	covariance
$\mathbb{E}$	expected value
ES	expected shortfall
$f(x \mid y)$	short for $f_{X\mid Y}(x \mid y)$; the density of X given Y
$\mathcal{F}_t$	information set up to time t, i.e., $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$
GARCH	generalized autoregressive conditional heteroskedastic
GEV	generalized extreme value
GPD	generalized Pareto distribution
$\mathbb{I}(A)$	indicator function; equal to 1 if $x \in A$, 0 otherwise
iid	independent and identically distributed
MDA	maximum domain of attraction
MEVD	multivariate extreme value distribution
MIDAS	mixed data sampling
MM	moving maxima
M3	maxima of moving maxima
M4	multivariate maxima of moving maxima
MCMC	Markov chain Monte Carlo
$\mathbb{N}$	set of natural numbers
Φ	standard Normal distribution function
Φ_α	Fréchet distribution function
ϕ	standard Normal density function
φ_α	Fréchet density function
$\mathbb{R}$	set of real numbers

SV	stochastic volatility
Var	variance
VaR	Value at Risk
$x_{s:t}$	the vector $(x_s, \ldots, x_t)'$, $s \leq t$
$\mathbb{Z}$	set of integer numbers

CHAPTER 1

Introduction

Intuitively, the objective of extreme value theory is to study probabilistic and statistical properties of very large or very small observations from random phenomenona. Multivariate extreme value theory is concerned with the joint distribution of extremes of multiple random variables. The theory is used in a number of areas such as finance and environmental science. For example, empirical observations suggest that extreme events in financial time series occur in clusters and are dependent across different assets.

We are interested in the extremal properties of multivariate stationary processes. Smith and Weissman (1996) showed that it is possible to characterize the extremal behavior of a multivariate stationary time series in terms of a limiting max-stable process. A key result of Smith and Weissman (1996) is that a large class of max-stable processes may be approximated by a particular class of max-stable processes known as *Multivariate Maxima of Moving Maxima processes* (or *M4 processes* for short). However, there has been little work on the statistical modeling of M4 processes. In this work we develop Monte Carlo methods for filtering, prediction, and parameter estimation of M4 processes.

1.1 Motivation

Some of the largest financial institutions have suffered losses in the hundreds of millions of dollars in the financial markets, mainly because of the lack of a good financial risk management system and the misuse of derivative products. This has increased the awareness of investors and regulators of the financial system to obtain accurate quantitative measures of financial risks. In particular, there is great concern about measuring market risk, which arises from fluctuations in the price of assets.

Value at Risk (VaR) has become the most widely accepted tool to measure market risk, and it is now a standard in the industry. Intuitively, VaR is the maximum loss that the value of an asset (or a portfolio of assets) can suffer with a given probability and during a specified time-horizon. In statistical terms the VaR can be thought of as a quantile of the returns distribution. An alternative measure of market risk with better theoretical properties is the *Expected Shortfall* (ES), which is the expected value of the loss given that the loss has exceeded the VaR. Both VaR and ES are properties of the tails of the distribution of asset returns, and a large literature in financial econometrics, and more recently in extreme value theory, has been devoted to the estimation of these quantities.

Empirical evidence suggests that financial time series have time-varying volatility and heavy tails. Much of the recent developments in financial econometrics have been models for time-varying volatility, such as ARCH and stochastic volatility models. However, volatility models are not as heavy-tailed as financial data and these models are fitted to the center of the distribution, so it is not clear that they also fit in the tails. Hence, the estimation of the tails of the distribution of asset returns is one of the main problems of current interest in financial econometrics. The need for accurate measures of financial market risk, such as VaR, has generated a great deal of attention to this problem.

Consider the standardized residuals from GARCH models fitted to four Nasdaq sector indices: Bank (BK), Industrial (ID), Insurance (IS), and Transportation (TR). Figure 1.1 shows exceedances of the standardized residuals above a large threshold, which illustrate two typical features of extremes of financial data. First, assuming that the standardized residuals are iid we would expect the exceedances to be scattered uniformly over the sample. However, there are periods where the exceedances in each individual series seem to occur in clusters, indicating that the extremes are dependent over time. Second, if there is an exceedance in one of the series on any given day it is likely that there is an exceedance in at least one of the other series on the same day, suggesting that there is joint dependence across the extremes of different assets. Hence there is a need for models that take into account temporal dependence as well as joint dependence in the extremes.

On the other hand, most multivariate extreme value models in the literature are limited to two or three dimensions. Thus, these models are not useful for typical portfolios, which

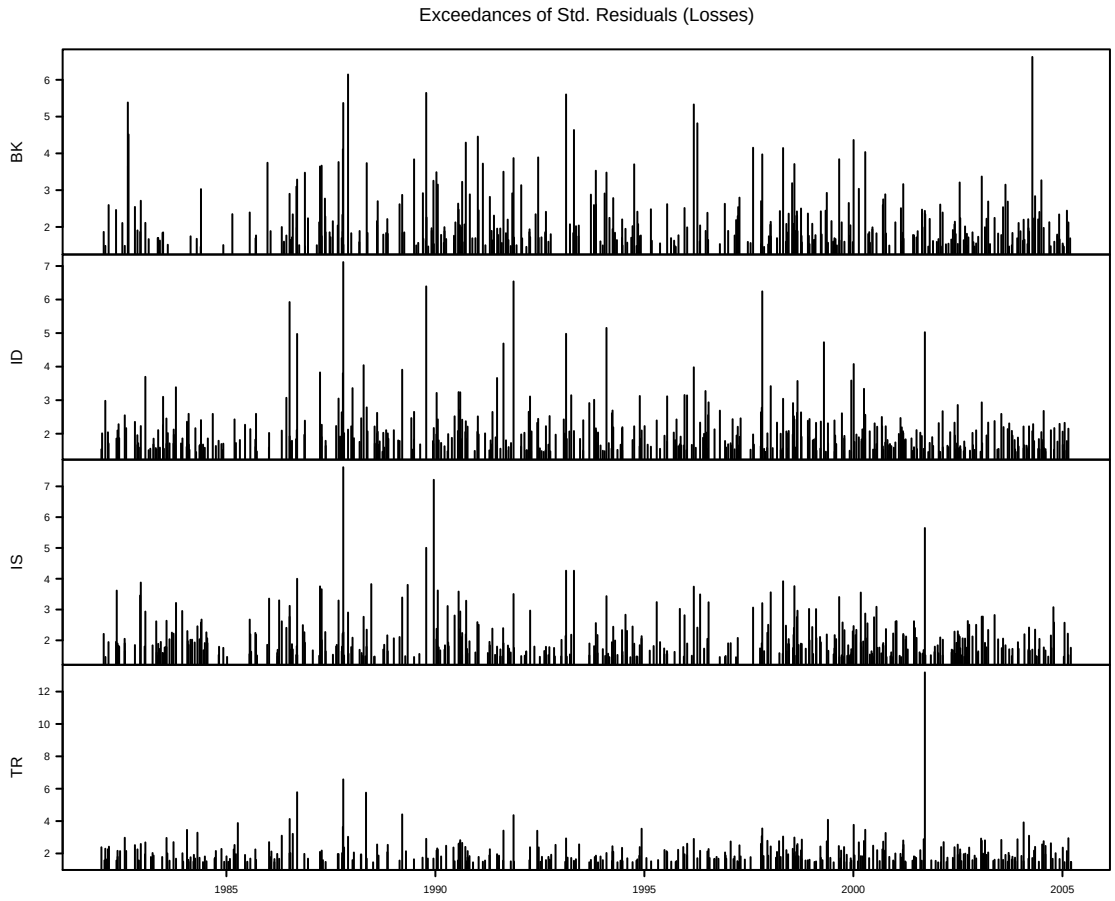

Figure 1.1: *Threshold Exceedances of GARCH Standardized Residuals.*

may include a large number of assets. A challenging problem is the estimation of measures of risk for the cumulative loss of a portfolio over a multiple-day period. For instance, the New Basel Capital Accord specifies that financial institutions should report VaR over a 10-day period[1].

Contribution

Following the work of Smith and Weissman (1996), Zhang (2002), and Smith (2003) we propose to model the joint extremal behavior of financial time series under the framework of extreme value theory for multivariate stationary processes. In particular, our approach for the statistical modeling of max-stable processes is based on *Moving Maxima* (MM) processes, and a multivariate extension known as *Multivariate Maxima of Moving Maxima*

[1] see http://www.bis.org/publ/bcbsca.htm

(M4) processes. The class of M4 processes allows to take into account clustering of extreme events and extremal dependence across assets that are empirically observed in the data.

One of the main problems in estimating the parameters of M4 processes is that the joint densities of M4 processes contain singularities because of the presence of deterministic signature patterns, so it is not possible to apply the method of maximum likelihood for parameter estimation. A solution to this problem is to write the M4 process as a state-space model, where the state is an unobserved M4 process, and the observed process is a nonlinear transformation of the state with small additive Gaussian noise to avoid degeneracies created by signature patterns. Our main contributions are as follows.

- In a state-space model it is usually assumed that the state is Markov. The theoretical results of this work show that two special cases of moving maxima processes, known as MM(1) and MM(2) processes, are second- and third-order Markov, respectively.

- Traditional estimation problems under a state-space framework are prediction, filtering, and smoothing. In this work we develop sequential Monte Carlo methods (particle filters) for prediction and filtering of M4 processes.

- Parameter estimation of M4 processes is one of the main challenges in applying these models to real data. In this work we develop Markov chain Monte Carlo algorithms for parameter estimation of moving maxima processes, which can be extended to M4 processes.

1.2 Outline

The rest of the document is organized as follows.

Chapter 2 provides the basic background for univariate and multivariate extreme value theory for iid observations, as well as the corresponding theory for stationary processes.

Chapter 3 introduces the class of moving maxima processes and its extensions, which include M3 and M4 processes. We review the literature on the estimation of M4 processes, and provide detailed algorithms for parameter estimation of M4 processes based on clustering methods, as originally proposed in Smith (2003).

Chapter 4 presents our theoretical results. We consider probabilistic properties of two special cases of moving maxima processes, which we refer to as MM(1) and MM(2) processes. Our main results show that MM(1) processes are second-order Markov, and MM(2) processes are third-order Markov.

Chapter 5 introduces the state-space approach for modeling time series, and provides the background for the class of simulation-based filters known as particle filters.

Chapter 6 is concerned with developing particle filtering methods for max-stable processes based on a state-space representation of M4 processes, assuming that the parameters are known. We start by developing specialized particle filtering methods for MM(1) processes based on the results from Chapter 4. We then propose particle filtering methods for general MM processes. We extend these methods to M3 processes, and finally we present the extensions to M4 processes.

Chapter 7 provides the background for Markov chain Monte Carlo (MCMC) methods, and develops detailed MCMC methods for the estimation of moving maxima processes. The extension of these methods for M4 processes is outlined.

Finally, Chapter 8 presents an overview of financial econometrics and financial risks. We discuss volatility models, and the definition of two measures of market risk: VaR and ES. We review other approaches that have been proposed for the estimation of VaR and ES. We also present our proposed solution to obtain one-step and multiple-step predictive distributions of M4 processes based on a simulation method. In the last section we apply our methods to the estimation of VaR and ES of Nasdaq sector indices over a multiple-day period.

CHAPTER 2

Extreme Value Theory

In this Chapter we review the basic background for univariate and multivariate extreme value theory for iid random variables, as well as the corresponding theory for stationary processes.

There is a rich literature on extreme value theory that goes back to the 1920's. Recent introductory books on the subject are Coles (2001), which emphasizes statistical modeling, and Finkenstädt and Rootzén (2003), which is a collection of accessible contributions with many applications of extreme value analysis. Embrechts, Klüppelberg and Mikosch (1997) is a comprehensive reference for the theory and its applications to insurance and finance. Beirlant, Goegebeur, Segers and Teugels (2004) provides a state-of-the-art review of the subject, combining theory and applications. Leadbetter, Lindgren and Rootzén (1983) is mostly concerned with extremes of stationary processes. For the theory of multivariate extremes we refer to Resnick (1987), Galambos (1987), and Kotz and Nadarajah (2000).

2.1 Univariate Extremes

This section summarizes the fundamental results in classical extreme value theory. We follow closely Embrechts *et al.* (1997) and Smith (2003).

2.1.1 Limit Laws for Maxima

Let $X_1, \ldots, X_n$ be iid random variables with common distribution function F, and let $M_n = \max\{X_1, \ldots, X_n\}$. Classical extreme value theory is concerned with the possible limit laws for M_n when properly normalized and centered. The main result is the *Fisher-Tippett Theorem*, also known as the *Three Types Theorem*.

Theorem 2.1.1 (Fisher-Tippett). *If there exist normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$, and some non-degenerate distribution function H such that for $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} P\left\{a_n^{-1}(M_n - b_n) \leq x\right\} = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x) \tag{2.1}$$

then H belongs to the type of one of the following three distribution functions:

$$\text{Gumbel:} \qquad \Lambda(x) = \exp\left\{-e^{-x}\right\}, \qquad x \in \mathbb{R} \tag{2.2}$$

$$\text{Fréchet:} \qquad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left\{-x^{-\alpha}\right\}, & x > 0, \end{cases} \qquad \alpha > 0 \tag{2.3}$$

$$\text{Weibull:} \qquad \Psi_\alpha(x) = \begin{cases} \exp\left\{-|x|^{\alpha}\right\}, & x \leq 0, \\ 1, & x > 0, \end{cases} \qquad \alpha > 0 \tag{2.4}$$

We refer to Leadbetter *et al.* (1983), and Resnick (1987) for a proof of Theorem 2.1.1.

The Gumbel, Fréchet, and Weibull distributions are also known as the *Extreme Value Distributions*. It is possible to combine the extreme value distributions into a single distribution, known as the *Generalized Extreme Value* (GEV) distribution. The GEV distribution with parameters $\xi \in \mathbb{R}$, $\psi > 0$, $\mu \in \mathbb{R}$, is defined by

$$H_{\xi,\psi,\mu}(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\psi}\right)_+^{-1/\xi}\right\}, \tag{2.5}$$

where $y_+ = \max(y, 0)$.

The GEV distribution with $\xi > 0$ corresponds to the Fréchet distribution with $\alpha = 1/\xi$; the limiting case $\xi \to 0$ corresponds to the Gumbel distribution; and the case $\xi < 0$ corresponds to the Weibull distribution with $\alpha = -1/\xi$. The parameters μ and ψ are location and scale parameters, respectively. The shape parameter ξ determines the tail behavior of the GEV distribution. An intuitive interpretation is that $\xi > 0$ corresponds to a long-tailed distribution; the limit $\xi \to 0$ to a distribution with exponential-type tail; and $\xi < 0$ to a short-tailed distribution with finite upper endpoint.

A non-degenerate distribution function F is said to be *max-stable* if for each $n = 2, 3, \ldots$

there are constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) = F(x). \tag{2.6}$$

The following result identifies the extreme value distributions with the class of max-stable distributions. This result is stated as Theorem 3.2.2 in Embrechts *et al.* (1997), and as Theorem 1.3.1(ii) in Leadbetter *et al.* (1983).

Theorem 2.1.2 (Limit Property of Max-Stable Laws). *The class of max-stable distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalized) maxima of iid random variables.*

A distribution function F belongs to the *maximum domain of attraction* of the extreme value distribution H, denoted $F \in \mathrm{MDA}(H)$, if there exist constants $a_n > 0,\ b_n \in \mathbb{R}$ such that for $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left\{a_n^{-1}(M_n - b_n) \le x\right\} = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x). \tag{2.7}$$

The theory of maximum domains of attraction is concerned with sufficient and necessary conditions for a distribution function F to satisfy (2.7), and characterizations of the normalizing constants a_n and b_n. For a comprehensive treatment of this subject we refer to Leadbetter *et al.* (1983), Resnick (1987), and Embrechts *et al.* (1997).

2.1.2 Limit Laws for Exceedances Over Thresholds

An alternative approach to study probabilistic properties of extremes of random phenomena is to consider exceedances over high thresholds. In particular this approach offers many advantages from a statistical point of view.

The excess distribution function over a threshold $u > 0$ is

$$F_u(x) = P\left\{X - u \le x \mid X > u\right\} = \frac{F(x+u) - F(u)}{1 - F(u)}, \tag{2.8}$$

where F is a distribution function with finite endpoint $x_F = \sup\{x : F(x) < 1\}$.

The *Generalized Pareto Distribution* (GPD) with parameters $\xi \in \mathbb{R}$, $\beta > 0$ is defined by

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi\frac{x}{\beta}\right)_+^{-1/\xi}. \tag{2.9}$$

The shape parameter ξ has the same interpretation as in the GEV distribution. The case $\xi > 0$ corresponds to a long-tailed distribution; the case $\xi < 0$ to a short-tailed distribution with finite upper endpoint; and the case $\xi = 0$ is interpreted as the limit when $\xi \to 0$, which is an exponential distribution $G_{0,\beta}(x) = 1 - \exp\{-x/\beta\}$.

The GPD appears as the limit distribution of scaled excesses over high thresholds, as shown in the following theorem, which basically states that $F_u(x) \to G_{\xi,\beta(u)}(x)$ for a large threshold u. (see Theorem 3.4.5 in Embrechts *et al.* (1997) and remarks that follow it).

Theorem 2.1.3. *Let X be a random variable with distribution function $F \in \mathrm{MDA}(H_{\xi,1,0})$, $\xi \in \mathbb{R}$. There exists a positive, measurable function $\beta(u)$ such that for $1 + \xi x > 0$,*

$$\lim_{u \uparrow x_F} P\left\{ \frac{X-u}{\beta(u)} > x \,\middle|\, X > u \right\} = (1+\xi x)_+^{-1/\xi}. \tag{2.10}$$

The *Poisson-GPD* model arises as a limiting form of the joint point process of exceedance times and excesses over the threshold. Let N be the number of exceedances of the level u in any one unit of time. In this model N has a Poisson distribution with mean λ, and conditionally on $N \geq 1$, the excess values $Y_1, \ldots, Y_N$ are iid with distribution function $G_{\xi,\beta}$. There is a relationship between the parameters of the GEV and the GPD models, which arises by calculating the distribution of the maximum exceedance in any one unit of time under the Poisson-GPD model (see also Theorem 3.4.13(d) in Embrechts *et al.*, 1997). If $x > u$,

$$\begin{aligned} P\left\{ \max_{1\leq i\leq N} Y_i \leq x \right\} &= P\{N=0\} + \sum_{n=1}^{\infty} P\{N=n,\ Y_1 \leq x, \ldots, Y_n \leq x\} \\ &= e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n} \left[1 - \left(1 + \xi\frac{x-u}{\beta}\right)_+^{-1/\xi}\right]^n \\ &= \exp\left\{ -\lambda \left(1 + \xi\frac{x-u}{\beta}\right)_+^{-1/\xi} \right\}. \end{aligned} \tag{2.11}$$

By substituting

$$\beta = \psi + \xi(u - \mu), \qquad \lambda = \left(1 + \xi\frac{u-\mu}{\psi}\right)^{-1/\xi}, \tag{2.12}$$

(2.11) reduces to the GEV form (2.5). Thus the GEV and GPD models are entirely consistent with one another above the threshold u, and (2.12) gives an explicit relationship between the two sets of parameters. In particular, the shape parameter ξ is the same in both models.

2.2 Multivariate Extremes

Multivariate extreme value theory is concerned with the joint distribution of extremes of multiple random variables. The probabilistic theory is now well developed and it is used in a number of areas such as finance and environmental science. Multivariate extremes are discussed in Galambos (1987), Resnick (1987), and Kotz and Nadarajah (2000), among others. In this section we follow closely Resnick (1987) and Smith, Tawn and Yuen (1990).

The first problem when dealing with multivariate extremes is that there is no natural extension of a notion of order in higher dimensions. The most widely used definition of maxima of a random vector is based on componentwise maxima.

In what follows we use boldface to denote vectors $\boldsymbol{x} = (x_1, \ldots, x_D)' \in \mathbb{R}^D$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid D-dimensional random vectors with common distribution function F, and define $M_{nd} = \max\{X_{1d}, \ldots, X_{nd}\}$ for each $d = 1, \ldots, D$. Suppose that for each $d = 1, \ldots, D$ there exist normalizing constants $a_{nd} > 0$, $b_{nd} \in \mathbb{R}$, and a D-dimensional distribution function H with non-degenerate marginals such that

$$\lim_{n\to\infty} P\left\{a_{nd}^{-1}\left(M_{nd} - b_{nd}\right) \leq x_d,\ 1 \leq d \leq D\right\}$$
$$= \lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nD}x_D + b_{nD}) = H(\boldsymbol{x}). \tag{2.13}$$

The possible limit distributions H are called the class of *Multivariate Extreme Value Distributions* (MEVDs). The main interest is in characterizing this class, but in contrast with the univariate case, there is no finite-dimensional parametric family that covers the whole class.

As in the univariate case, we say that $F \in \text{MDA}(H)$ if for each $d = 1, \ldots, D$ there exist constants $a_{nd} > 0,\ b_{nd} \in \mathbb{R}$ such that (2.13) is satisfied.

In a multivariate context, a distribution function F is *max-stable* if for each $d = 1, \ldots, D$, and for each $n = 2, 3, \ldots$ there are constants $a_{nd} > 0$ and $b_{nd} \in \mathbb{R}$ such that

$$F^n(a_{n1}\, x_1 + b_{n1}, \ldots, a_{nD}\, x_D + b_{nD}) = F(\boldsymbol{x}). \tag{2.14}$$

Parallel to the univariate case, the class of MEVDs is characterized by a max-stability property. The following is Proposition 5.9 in Resnick (1987).

Theorem 2.2.1. *The class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with non-degenerate marginals.*

From the univariate theory for extremes, it follows that the marginals of the limiting distribution H must be a one-dimensional extreme value distribution, but in order to characterize max-stable distributions it is helpful to standardize the marginals. Resnick (1987) mentions that different marginal assumptions have led to different representations, but in the end they are essentially equivalent. Here we assume the marginals of H are unit Fréchet. There is no loss of generality in assuming unit Fréchet marginals because we can always apply a transformation to achieve this. Note also that in this case, we may take the normalizing constants to be $a_{nd} = n,\ b_{nd} = 0$, for each $d = 1, \ldots, D$.

Under this framework any limiting distribution of normalized componentwise maxima, with unit Fréchet marginals, can be characterized as follows.

Theorem 2.2.2. *$H(\boldsymbol{x})$ is a multivariate extreme value distribution with unit Fréchet marginals if and only if there exists a finite measure S on $\aleph = \{\boldsymbol{x} \in E : \|\boldsymbol{x}\| = 1\}$, where $E = [0, \infty]^D \setminus \{\mathbf{0}\}$, satisfying*

$$\int_{\aleph} w_d \, \mathrm{d}S(\boldsymbol{w}) = 1, \qquad d = 1, \ldots, D \tag{2.15}$$

such that for $\boldsymbol{x} \in \mathbb{R}^D$

$$H(\boldsymbol{x}) = \exp\left\{ - \int_{\aleph} \max_{1 \le d \le D} \frac{w_d}{x_d} \, \mathrm{d}S(\boldsymbol{w}) \right\}. \tag{2.16}$$

Galambos (1987) reproduces J. Pickands' proof of Theorem 2.2.2. A different proof can be found in Proposition 5.11 in Resnick (1987), which is a more general version of Theorem 2.2.2. Resnick (1987) also discusses in detail other characterizations and properties of MEVDs, and it also presents a more general characterization based on a point process representation. For a recent review on multivariate extremes we refer to Fougères (2003).

2.3 Extremes of Stationary Processes

So far we have discussed extreme properties for iid random variables. Many applications require modeling temporal dependence of extremes, so a natural extension is to analyze the extremal behavior of stationary stochastic processes. In studying the extremes of temporal-dependent sequences it is usually assumed there is asymptotic independence between observations with large time-separation. This idea allows to relate the extremes of stationary processes to those of iid sequences, by means of a key parameter known as the *extremal index*. In this section we follow closely Leadbetter *et al.* (1983), Smith and Weissman (1996), and Smith (2003).

2.3.1 The Extremal Index

Let $\{X_t,\ t = 1, 2, \ldots\}$ be a discrete-time strictly stationary stochastic process with marginal distribution function F and let $M_n = \max\{X_1, \ldots, X_n\}$. Following Leadbetter *et al.* (1983) we define the extremal index as follows. The process $\{X_t\}$ has *extremal index* $\theta \in [0, 1]$ if for each $\tau > 0$ there exists a sequence $\{u_n\}$ such that

$$\lim_{n\to\infty} n\big(1 - F(u_n)\big) = \tau, \tag{2.17}$$

and

$$\lim_{n\to\infty} P\{M_n \leq u_n\} = e^{-\theta\tau}. \tag{2.18}$$

The extremal index does not depend on τ or the sequence $\{u_n\}$, and it is a constant for the process. Intuitively, the extremal index can be thought of as the inverse of the mean cluster size in the point process of exceedance times over a high threshold. We refer to Leadbetter *et al.* (1983) for a rigorous development of extreme value theory for stationary

processes.

Let $\widetilde{X}_1, \widetilde{X}_2, \ldots$ be the so-called *associated sequence* of iid random variables with the same marginal distribution function F as the process $\{X_t\}$, and let $\widetilde{M}_n = \max\{\widetilde{X}_1, \ldots, \widetilde{X}_n\}$. The following result (Leadbetter *et al.*, 1983, Theorem 3.7.2(i)) illustrates the relationship between the process $\{X_t\}$ with extremal index θ, and the associated iid sequence $\{\widetilde{X}_t\}$.

Theorem 2.3.1. *Suppose that* (2.17) *holds, then*

$$\lim_{n\to\infty} P\{\widetilde{M}_n \leq u_n\} = e^{-\tau} \quad \textit{if and only if} \quad \lim_{n\to\infty} P\{M_n \leq u_n\} = e^{-\theta\tau}. \tag{2.19}$$

For multivariate processes the parameter of interest is the *multivariate extremal index*, originally proposed by Nandagopalan (1990, 1994). Following Smith and Weissman (1996), let $\{\boldsymbol{X}_t,\ t = 1, 2, \ldots\}$ be a D-dimensional stationary process with marginal distribution functions $F(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^D$, and $F_d(x),\ d = 1, \ldots, D$. Denote the vector of componentwise maxima by $\boldsymbol{M}_n = (M_{n1}, \ldots, M_{nD})$. Let $\{\widetilde{\boldsymbol{X}}_t,\ t = 1, 2, \ldots\}$ be the associated sequence of D-dimensional iid random vectors with the same distribution function F, and let $\widetilde{\boldsymbol{M}}_n$ denote the corresponding vector of componentwise maxima. Suppose $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_D)$ is a vector of nonnegative finite numbers, and suppose for each $d = 1, \ldots, D$, $\{u_{nd}\}$ is a sequence of thresholds such that

$$\lim_{n\to\infty} n\big(1 - F_d(u_{nd})\big) = \tau_d. \tag{2.20}$$

Now consider the following joint limits

$$\lim_{n\to\infty} P\{M_{nd} \leq u_{nd},\ 1 \leq d \leq D\} = H(\boldsymbol{\tau}), \tag{2.21}$$

$$\lim_{n\to\infty} P\{\widetilde{M}_{nd} \leq u_{nd},\ 1 \leq d \leq D\} = \widetilde{H}(\boldsymbol{\tau}). \tag{2.22}$$

If both (2.21) and (2.22) exist and are non-zero, then the *multivariate extremal index* $\theta(\boldsymbol{\tau})$ is defined by

$$\theta(\boldsymbol{\tau}) = \frac{\log H(\boldsymbol{\tau})}{\log \widetilde{H}(\boldsymbol{\tau})}. \tag{2.23}$$

The multivariate extremal index satisfies $0 \leq \theta(\boldsymbol{\tau}) \leq 1$ for all $\boldsymbol{\tau}$, and it has the property $\theta(c\boldsymbol{\tau}) = \theta(\boldsymbol{\tau})$ for any $c > 0$. However, the multivariate extremal index is a function of $\boldsymbol{\tau}$,

so in contrast to the univariate case it is not constant.

2.3.2 Max-Stable Processes

Following the development of multivariate extremes, it is sufficient to consider processes with unit Fréchet marginals. A process $\{X_t\}$ with unit Fréchet marginals is called *max-stable* if all its finite-dimensional distributions are max-stable, i.e. for any $n \in \mathbb{N},\ r \in \mathbb{N}$,

$$P\{X_t \leq nx_t,\ 1 \leq t \leq r\}^n = P\{X_t \leq x_t,\ 1 \leq t \leq r\}. \tag{2.24}$$

Similarly, a D-dimensional process $\{\boldsymbol{X}_t\}$ with unit Fréchet marginals is max-stable if for any $n \in \mathbb{N},\ r \in \mathbb{N}$,

$$P\{X_{td} \leq nx_{td},\ 1 \leq t \leq r, 1 \leq d \leq D\}^n = P\{X_{td} \leq x_{td},\ 1 \leq t \leq r, 1 \leq d \leq D\}. \tag{2.25}$$

A process $\{\boldsymbol{Y}_t\}$ is said to be in the domain of attraction of a max-stable process $\{\boldsymbol{X}_t\}$ if there exist normalizing constants $a_{ntd} > 0,\ b_{ntd} \in \mathbb{R}$ such that for any finite $r \in \mathbb{N}$,

$$\lim_{n\to\infty} P\left\{a_{ntd}^{-1}(Y_{td} - b_{ntd}) \leq x_{td},\ 1 \leq t \leq r, 1 \leq d \leq D\right\}^n = P\{X_{td} \leq x_{td},\ 1 \leq t \leq r, 1 \leq d \leq D\}. \tag{2.26}$$

Smith and Weissman (1996) made a connection between max-stable processes and the limiting distributions of extreme values in dependent stochastic processes. This result is Theorem 2.3 of Smith and Weissman (1996), which states that if (2.26) holds, together with some mixing conditions on $\{\boldsymbol{Y}_t\}$ and $\{\boldsymbol{X}_t\}$, then both processes have the same multivariate extremal index. Since the multivariate extremal index essentially captures what we need to know about the extremes of multivariate stationary processes, this result shows that we can learn about the extremal behavior of a multivariate stationary time series if we look at the limiting max-stable process. In the following Chapter we consider a particular class of max-stable processes.

CHAPTER 3

Representations of Max-Stable Processes

In this Chapter we introduce a particular class of max-stable processes known as *Moving Maxima* (MM) processes, and a multivariate extension known as *Multivariate Maxima of Moving Maxima* (M4) processes. As previously mentioned, Smith and Weissman (1996) showed that the extremes of multivariate stationary processes can be characterized in terms of a limiting max-stable process. Furthermore, Smith and Weissman (1996) showed that a large class of max-stable processes may be approximated by M4 processes. Thus we consider M4 processes as the main object of study.

In this Chapter we also discuss a feature of these models called *signature patterns*, and we review the literature on the estimation of M4 processes. In particular, we provide full details of an estimation method originally proposed in Smith (2003), which is based on clustering methods.

3.1 Moving Maxima (MM) Processes

Let $\{\alpha_k,\ k \in \mathbb{Z}\}$ be a sequence of nonnegative constants satisfying $\sum_k \alpha_k = 1$, and let $\{Z_t,\ t \in \mathbb{Z}\}$ be a sequence of iid unit Fréchet random variables. The *Moving Maximum* process (MM process for short) is defined by

$$X_t = \max_{k \in \mathbb{Z}} \alpha_k Z_{t-k}, \qquad t \in \mathbb{Z}. \tag{3.1}$$

The finite-dimensional distributions of these processes are, for any $r \in \mathbb{N}$,

$$P\{X_t \leq x_t,\ 1 \leq t \leq r\} = \exp\left\{ -\sum_{m=-\infty}^{\infty} \max_{1-m \leq k \leq r-m} \frac{\alpha_k}{x_{m+k}} \right\}. \tag{3.2}$$

It is easily seen from (3.2) that $\{X_t\}$ is a stationary process with unit Fréchet marginals. Moreover, (3.2) satisfies (2.24), thus MM processes are max-stable. It can be verified (for instance, Example 10.5 in Beirlant *et al.*, 2004) that the extremal index of MM processes is

$$\theta = \max_k \alpha_k. \tag{3.3}$$

Example 3.1.1 (MM(1) Processes). Let $\alpha \in (0,1)$, and let $\{Z_t\}$ be a sequence of iid unit Fréchet random variables. Define

$$X_t = \max\{\alpha Z_t,\, (1-\alpha)Z_{t-1}\}, \qquad t \in \mathbb{Z}. \tag{3.4}$$

For convenience we refer to processes defined by (3.4) as *MM(1) processes*. The extremal index of MM(1) processes is

$$\theta = \max\{\alpha, 1-\alpha\} > \frac{1}{2}. \tag{3.5}$$

Hence, for these processes $1/2 \leq \theta \leq 1$. The interpretation is that if $\theta \to 1$ then we have independence between contiguous observations, and the mean cluster size is 1. If $\theta \to 1/2$ then we have perfect dependence because for large X_t there is high probability that $X_{t+1} = X_t$, so the mean cluster size is 2. Simulated sample paths of MM(1) processes with $\alpha = 0.5, 0.7$ are shown in Figure 3.1. ‖

3.2 Maxima of Moving Maxima (M3) Processes

Deheuvels (1983) introduced a more general class of max-stable processes based on superpositions of independent MM processes. Let $\{\alpha_{\ell k},\ \ell \in \mathbb{N},\, k \in \mathbb{Z}\}$ be a double sequence of nonnegative constants satisfying $\sum_\ell \sum_k \alpha_{\ell k} = 1$, and let $\{Z_{\ell t},\ \ell \in \mathbb{N},\, t \in \mathbb{Z}\}$ be a double sequence of iid unit Fréchet random variables. The *Maxima of Moving Maxima* process (M3 process for short) is defined by

$$X_t = \max_{\ell \in \mathbb{N}} \max_{k \in \mathbb{Z}} \alpha_{\ell k} Z_{\ell, t-k}, \qquad t \in \mathbb{Z}. \tag{3.6}$$

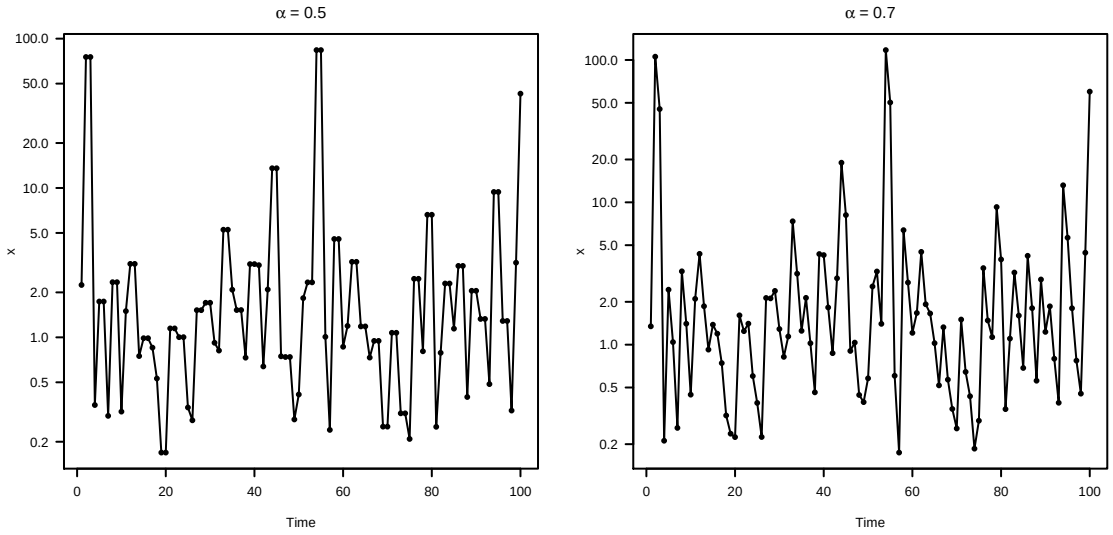

Figure 3.1: *Sample Paths of Simulated MM(1) Processes:* $\alpha = 0.5$ *(left), and* $\alpha = 0.7$ *(right).*

The finite-dimensional distributions of these processes are, for any $r \in \mathbb{N}$,

$$P\left\{X_t \leq x_t,\ 1 \leq t \leq r\right\} = \exp\left\{-\sum_{\ell=1}^{\infty}\sum_{m=-\infty}^{\infty}\max_{1-m\leq k\leq r-m}\frac{\alpha_{\ell k}}{x_{m+k}}\right\}. \tag{3.7}$$

It can be verified from (3.7) that M3 processes are max-stable stationary processes with unit Fréchet marginals. The extremal index of M3 processes is

$$\theta = \sum_{\ell}\max_{k}\alpha_{\ell k}. \tag{3.8}$$

Remark. One can think of an M3 process as a superposition of independent MM processes. For instance, let $X_t^{(1)}$ and $X_t^{(2)}$ be two independent MM processes and consider the process

$$\begin{aligned} Y_t &= \max\left\{X_t^{(1)},\ X_t^{(2)}\right\} \\ &= \max\left\{\max_k \alpha_{1,k}Z_{1,t-k},\ \max_j \alpha_{2,j}Z_{2,t-j}\right\} \\ &= \max_{\ell=1,2}\max_k \alpha_{\ell,k}Z_{\ell,t-k}, \qquad t \in \mathbb{Z}. \end{aligned} \tag{3.9}$$

Then Y_t is an M3 process.

3.3 Multivariate Maxima of Moving Maxima (M4) Processes

Smith and Weissman (1996) introduced a class of multivariate max-stable processes that are a generalization of M3 processes. These processes are called *Multivariate Maxima of Moving Maxima* processes (M4 processes for short), and are defined as follows. Let $\{\alpha_{\ell kd},\ \ell \in \mathbb{N},\ k \in \mathbb{Z},\ 1 \le d \le D\}$ be a triple sequence of nonnegative constants satisfying

$$\sum_{\ell} \sum_{k} \alpha_{\ell kd} = 1, \qquad \text{for each } d = 1, \ldots, D, \tag{3.10}$$

and let $\{Z_{\ell t},\ \ell \in \mathbb{N},\ t \in \mathbb{Z}\}$ be a double sequence of iid unit Fréchet random variables. The M4 process is defined by

$$X_{td} = \max_{\ell \in \mathbb{N}} \max_{k \in \mathbb{Z}} \alpha_{\ell kd} Z_{\ell, t-k}, \qquad t \in \mathbb{Z}, \quad 1 \le d \le D. \tag{3.11}$$

Smith and Weissman (1996) derived the finite-dimensional distributions of these processes as follows. For any $r \in \mathbb{N}$,

$$\begin{aligned}
&P\left\{X_{td} \le x_{td}, \quad 1 \le t \le r,\ 1 \le d \le D\right\} \\
&= P\left\{Z_{\ell,t-k} \le \frac{x_{td}}{\alpha_{\ell kd}}, \quad \ell \in \mathbb{N},\ k \in \mathbb{Z},\ 1 \le t \le r,\ 1 \le d \le D\right\} \\
&= P\left\{Z_{\ell m} \le \min_{1-m \le k \le r-m} \min_{1 \le d \le D} \frac{x_{m+k,d}}{\alpha_{\ell kd}}, \quad \ell \in \mathbb{N},\ m \in \mathbb{Z}\right\} \\
&= \exp\left\{-\sum_{\ell=1}^{\infty} \sum_{m=-\infty}^{\infty} \max_{1-m \le k \le r-m} \max_{1 \le d \le D} \frac{\alpha_{\ell kd}}{x_{m+k,d}}\right\}.
\end{aligned} \tag{3.12}$$

It can be verified from (3.12) that M4 processes are max-stable stationary processes with unit Fréchet marginals. Moreover, Smith and Weissman (1996) showed that the multivariate extremal index is,

$$\theta(\boldsymbol{\tau}) = \frac{\sum_{\ell} \max_k \max_d \alpha_{\ell kd} \tau_d}{\sum_{\ell} \sum_k \max_d \alpha_{\ell kd} \tau_d}, \tag{3.13}$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_D)$ is a vector of nonnegative finite numbers.

Example 3.3.1. Figure 3.2 shows $T = 500$ observations from a simulated M4 process in $D = 3$ dimensions with parameters given in Table 3.1. The observations around $t = 165$

illustrate typical behavior of M4 processes: extreme values tend to occur in clusters and are dependent across series. ||

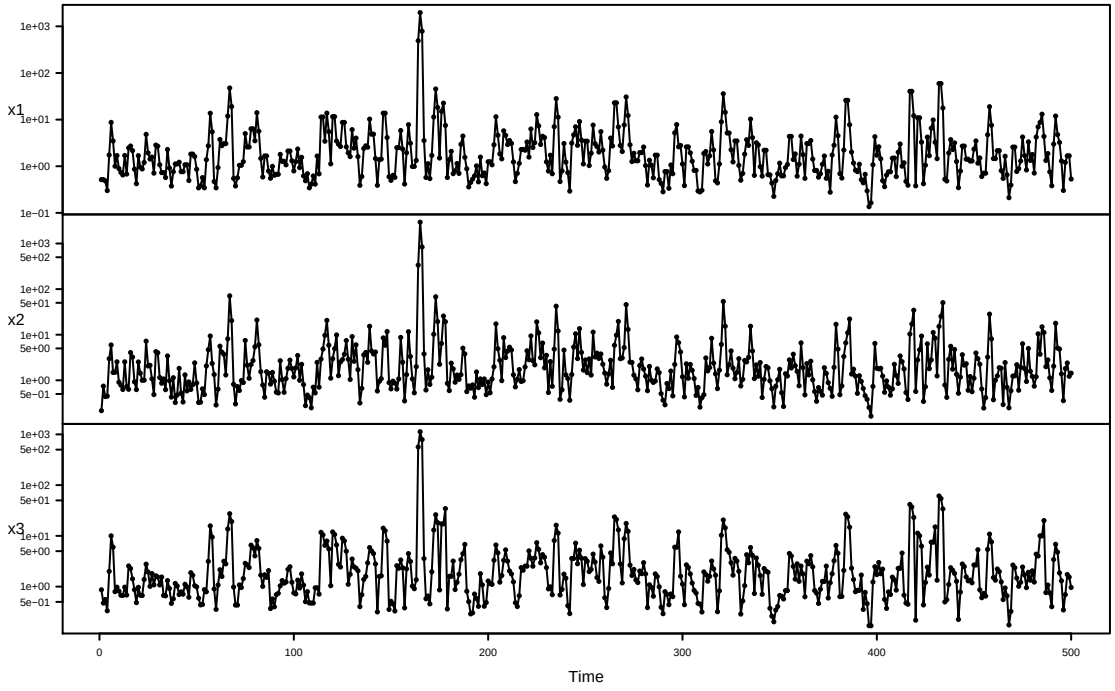

Figure 3.2: *Sample Path of Simulated M4 Process.*

3.4 Signature Patterns

One of the main features of M4 processes is that a single large shock observed in the process creates a deterministic *signature pattern*. We illustrate this concept with MM(1) processes and then define it for M4 processes.

Consider the MM(1) process $X_t = \max\{\alpha Z_t, (1-\alpha)Z_{t-1}\}$. Suppose that for some t^* the value of Z_{t^*} is much larger than its neighbors Z_{t^*-1} and Z_{t^*+1}, such that

$$X_{t^*} = \alpha Z_{t^*} \quad \text{and} \quad X_{t^*+1} = (1-\alpha)Z_{t^*}. \tag{3.14}$$

Then the signature pattern of this process is defined as the ratio

$$S_{t^*+1} = \frac{X_{t^*+1}}{X_{t^*}} = \frac{1-\alpha}{\alpha}. \tag{3.15}$$

ℓ	$\alpha_{\ell,-1}$	$\alpha_{\ell,0}$	$\alpha_{\ell,1}$
$d=1$			
1	0.0714	0.2857	0.1143
2	0.1429	0.1429	0.0429
3	0.0143	0.0714	0.0286
4	0.0286	0.0429	0.0143
$d=2$			
1	0.0488	0.4268	0.1220
2	0.0366	0.0610	0.1220
3	0.0244	0.0488	0.0122
4	0.0122	0.0488	0.0366
$d=3$			
1	0.0820	0.1639	0.1148
2	0.1475	0.1311	0.0820
3	0.0164	0.0820	0.0492
4	0.0328	0.0328	0.0656

Table 3.1: *Parameters of Simulated M4 Process. We assume* $\alpha_{\ell,k,d} = 0$ *for indices* $(\ell, k, d) \in \{(\ell, k, d) : |k| > 1,\ \ell > 4,\ 1 \leq d \leq 3\}$.

Solving for α we obtain

$$\alpha = \frac{X_{t^*}}{X_{t^*+1} + X_{t^*}}. \tag{3.16}$$

Zhang (2002) showed that (3.16) will hold infinitely often, so if we observe the process for a long period of time the relationship (3.16) creates a deterministic pattern that identifies the parameter of the process.

The distribution of X_{t^*+1} is a mixture of a continuous random variable (Z_{t^*+1}), and a degenerate random variable that takes the value $\frac{1-\alpha}{\alpha} x_{t^*}$ with probability 1. In principle, the relationship (3.16) can be used to construct a maximum likelihood approach for the estimation of the parameter α as follows (see also Hall *et al.*, 2002). If two of the ratios X_t/X_{t+1} and X_s/X_{s+1}, $s \neq t$ are identical, then the value of α can be identified exactly. However, in practice we cannot expect an actual time series to follow this model exactly, so the ratio (3.16) will never give the exact value of α. Therefore, a maximum likelihood approach is not feasible.

The M4 processes defined by (3.11) have indices ℓ and k that range over an infinite (countable) set. In practice it is necessary to restrict the indices ℓ and k to range over a finite index set $\{(k, \ell) : 1 \leq \ell \leq L,\ -K_1 \leq k \leq K_2\}$, where L, K_1 and K_2 are known positive integers. To simplify our notation, and without loss of generality, from here on we

assume $K = K_1 = K_2$. Hence the process we consider is

$$X_{td} = \max_{1\le\ell\le L} \max_{|k|\le K} \alpha_{\ell kd} Z_{\ell,t-k}, \qquad t \in \mathbb{N}, \quad 1 \le d \le D, \tag{3.17}$$

where $\sum_\ell \sum_k \alpha_{\ell kd} = 1$ for each $d = 1, \ldots, D$.

Following the argument in Smith (2003), consider a situation where some value of $\{Z_{\ell,t}\}$, say Z_{ℓ^*,t^*}, is much larger than its neighbors $\{Z_{\ell,t^*+k},\ 1 \le \ell \le L,\ |k| \le K\}$. Then we have

$$X_{t^*+k,d} = \alpha_{\ell^*,k,d} Z_{\ell^*,t^*}, \qquad |k| \le K, \quad 1 \le d \le D. \tag{3.18}$$

Thus the ℓth *signature pattern* is defined by

$$S_{t^*+k,d} = \frac{X_{t^*+k,d}}{\max_k \max_d X_{t^*+k,d}} \tag{3.19}$$

$$= \frac{\alpha_{\ell^*,k,d}}{\max_k \max_d \alpha_{\ell^*,k,d}}, \qquad |k| \le K, \quad 1 \le d \le D. \tag{3.20}$$

A signature pattern specifies the shape of the process near its local maximum. For any (ℓ^*, t^*), there is positive probability that (3.19) holds exactly. There are L such deterministic signature patterns, and Zhang and Smith (2003) have shown rigorously that each of the L signature patterns will occur infinitely often. This means that the joint densities of M4 processes contain singularities because of the presence of these deterministic signature patterns, so it is not possible to apply the method of maximum likelihood to estimate the parameters of the model. Therefore alternative estimation methods are required.

3.5 Literature Review on the Estimation of M4 Processes

The results of Smith and Weissman (1996) allow us to characterize the extremal behavior of a multivariate stationary time series in terms of a limiting max-stable process. However, there has been little work on the statistical modeling of max-stable processes. The key characterization result of Smith and Weissman (1996) is that any D-dimensional stationary max-stable process with unit Fréchet marginals may be approximated arbitrarily closely as the sum of an M4 process and a deterministic process—a generalization of a result of Deheuvels (1983) for one-dimensional processes. It is usually assumed that the

deterministic process is absent, so the practical interpretation is that we can approximate any max-stable process with unit Fréchet marginals by an M4 process. Hence the problem of estimating max-stable processes can be reduced to the estimation of M4 processes.

A special class of moving maxima processes called *max-autoregressive moving average* (MARMA) processes was introduced in Davis and Resnick (1989). A stationary process $\{X_t\}$ is a MARMA(p, q) process if it satisfies

$$X_t = \max\left\{\phi_1 X_{t-1}, \ldots, \phi_p X_{t-p}, Z_t, \theta_1 Z_{t-1}, \ldots, \theta_q Z_{t-q}\right\}, \quad t \in \mathbb{N}, \tag{3.21}$$

where $\phi_i \geq 0,\ i = 1, \ldots, p,\ \theta_j \geq 0,\ j = 1, \ldots, q$, and $\{Z_t\}$ are iid unit Fréchet random variables. Davis and Resnick (1989) only provided an estimation method for the special case when $q = 1$.

Hall, Peng and Yao (2002) considered moving maxima processes and constructed confidence and prediction intervals based on empirical distributions and bootstrap techniques.

Zhang (2002) generalized the approach of Hall *et al.* (2002) to M4 processes, but without the use of bootstrap techniques. Zhang (2002) proposed a series of estimating procedures based on identifying signature patterns, and showed consistency and asymptotic normality of the parameter estimators.

In practice it is unrealistic to assume that we can observe the exact signature patterns that are characteristic of M4 processes. The approach of Smith (2003) is to define candidate signature patterns on blocks of observations, and then use a clustering algorithm to identify each of the L signature patterns defined by (3.20). However, a precise description of how to obtain the parameter estimates using this approach has not been discussed elsewhere.

3.6 Clustering-Based Estimation of M4 Processes

In the following sections we provide a detailed description of the method proposed by Smith (2003) for parameter estimation of M4 processes. To motivate this approach consider the MM(1) process $X_t = \max\{\alpha Z_t, (1-\alpha)Z_{t-1}\}$, and suppose Z_{t^*} is much larger than its neighbors Z_{t^*-1} and Z_{t^*+1}, so that (3.16) holds. Zhang (2002) showed that (3.16) will hold infinitely often, so in principle we only need two very large observations at different times

t_1, t_2 in order to identify the exact value of α via (3.16). However, in practice we cannot expect for any time series to follow this model exactly, so we proceed as follows. Fix a large threshold $u > 0$ and consider a moving window of size 2. Then for $t = 1, \ldots, T-1$, if $x_t > \max\{u, x_{t+1}\}$ calculate a *candidate* signature pattern

$$a^{(t)} = \frac{x_t}{x_{t+1} + x_t}. \tag{3.22}$$

Let n be the sample size of the set $\left\{a^{(t_i)}\right\}$, then the sample mean of the candidate signature patterns gives an estimate of α:

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} a^{(t_i)}. \tag{3.23}$$

This idea can be easily generalized for the estimation of more general MM processes, as described in the following section.

3.6.1 Estimation of MM Processes

Consider the MM process

$$X_t = \max_{|k| \leq K} \alpha_k Z_{t-k}, \qquad t \in \mathbb{N}, \tag{3.24}$$

where $\sum_k \alpha_k = 1$.

Suppose Z_{t^*} is much larger than its neighbors, so that

$$X_{t^*+k} = \alpha_k Z_{t^*}, \qquad -K \leq k \leq K. \tag{3.25}$$

Then the signature pattern $\boldsymbol{S}_{t^*} = (S_{t^*-K}, \ldots, S_{t^*-1}, S_{t^*}, S_{t^*+1}, \ldots, S_{t^*+K})'$ is defined by

$$S_{t^*+k} = \frac{X_{t^*+k}}{X_{t^*}} = \frac{\alpha_k}{\alpha_0}, \qquad -K \leq k \leq K. \tag{3.26}$$

The restriction $\sum_k \alpha_k = 1$ gives

$$\sum_{k=-K}^{K} S_{t^*+k} = \frac{\sum_{k=-K}^{K} X_{t^*+k}}{X_{t^*}} = \frac{1}{\alpha_0}. \tag{3.27}$$

Defining

$$X_t^+ = \sum_{k=-K}^{K} X_{t+k}, \tag{3.28}$$

and solving for α_0 we have

$$\alpha_0 = \frac{X_{t^*}}{X_{t^*}^+}. \tag{3.29}$$

Substituting (3.29) into (3.26) we obtain

$$\begin{aligned} \alpha_k &= \alpha_0 \frac{X_{t^*+k}}{X_{t^*}} \\ &= \frac{X_{t^*+k}}{X_{t^*}^+}, \qquad -K \le k \le K. \end{aligned} \tag{3.30}$$

Zhang (2002) showed that (3.30) will hold infinitely often. However, in practice we cannot expect the actual data to follow this model exactly. This problem is solved by defining candidate signature patterns for each *local maximum*.

Definition 3.6.1 (Local Maximum). For a given integer K and a positive threshold u, consider a block of observations $x_{t-K:t+K} = (x_{t-K}, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_{t+K})$, $K < t < T - K$. We say that x_t is a *local maximum* if $x_t \ge \max\{u, x_{t-K:t+K}\}$.

Remark. Without loss of generality, we define the local maximum to be at the center of the block. This means that we are assuming that $\alpha_0 = \max_k \alpha_k$.

The method to estimate the parameters of MM process is summarized in Algorithm 3.6.2.

Algorithm 3.6.2 (Clustering-Based Estimation of MM Processes).

1. Consider a moving window of size $2K + 1$, and fix a large threshold $u > 0$.
2. For $t = K + 1, \ldots, T - K$, if $x_t > \max\{u, x_{t-K:t+K}\}$ then calculate

$$a_k^{(t)} = \frac{x_{t+k}}{x_t^+}, \qquad -K \le k \le K. \tag{3.31}$$

3. Calculate

$$\hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^{n} a_k^{(t_i)}, \qquad -K \le k \le K. \tag{3.32}$$

where n is the number of local maxima.

Example 3.6.3 (MMA(q) Processes). Consider the special case of a *max-moving average* MMA(q) process (Davis and Resnick, 1989),

$$X_t = \max\{Z_t, \theta_1 Z_{t-1}, \ldots, \theta_q Z_{t-q}\}, \qquad t \in \mathbb{N}, \tag{3.33}$$

where $\theta_i > 0,\ i = 1, \ldots, q$ and $\{Z_i\}$ are iid random variables. The values of θ_i do not have to add up to 1, so we set $\theta_0 = 1$, $C = \sum_{i=0}^{q} \theta_i$, and write

$$X_t = C \max\{\tilde{\alpha}_0 Z_t, \tilde{\alpha}_1 Z_{t-1}, \ldots, \tilde{\alpha}_q Z_{t-q}\}, \qquad t \in \mathbb{N}, \tag{3.34}$$

where $\tilde{\alpha}_i = C^{-1}\theta_i,\ i = 0, \ldots, q$. Using this parameterization we can use Algorithm 3.6.2 to obtain estimates of

$$\tilde{\alpha}_0 = C^{-1} \tag{3.35}$$

$$\tilde{\alpha}_i = \theta_i\, \tilde{\alpha}_0, \qquad i = 1, \ldots, q. \tag{3.36}$$

The estimates of the original parameters are $\theta_i = \tilde{\alpha}_i/\tilde{\alpha}_0,\ i = 1, \ldots, q$. ‖

3.6.2 Estimation of M3 Processes

Consider the M3 process

$$X_t = \max_{1 \le \ell \le L} \max_{|k| \le K} \alpha_{\ell k} Z_{\ell, t-k}, \qquad t \in \mathbb{N}, \tag{3.37}$$

where $\sum_\ell \sum_k \alpha_{\ell k} = 1$.

The M3 process we consider has L signature patterns indexed by ℓ. Smith and Weissman (1996) pointed out that the relative frequency of the ℓth signature pattern is

$$\alpha_\ell^+ = \sum_{k=-K}^{K} \alpha_{\ell k}. \tag{3.38}$$

We can interpret that in the long run, α_ℓ^+ is the proportion of time that the process is driven by the ℓth signature pattern. This is the key feature of the model used in our estimation procedure.

Suppose Z_{ℓ^*,t^*} is much larger than its neighbors, so there is a single index (ℓ^*, t^*), such that

$$X_{t^*+k} = \alpha_{\ell^*,k} Z_{\ell^*,t^*}, \qquad -K \le k \le K. \tag{3.39}$$

Then we have

$$S_{\ell^*,t^*+k} = \frac{X_{t^*+k}}{X_{t^*}} = \frac{\alpha_{\ell^*,k}}{\alpha_{\ell^*,0}}, \qquad -K \le k \le K, \tag{3.40}$$

which defines the ℓ^*th signature pattern

$$\boldsymbol{S}_{\ell^*,t^*} = (S_{\ell^*,t^*-K}, \ldots, S_{\ell^*,t^*-1}, S_{\ell^*,t^*}, S_{\ell^*,t^*+1}, \ldots, S_{\ell^*,t^*+K})'. \tag{3.41}$$

Note that

$$\sum_{k=-K}^{K} S_{\ell^*,t^*+k} = \frac{\sum_{k=-K}^{K} X_{t^*+k}}{X_{t^*}} = \frac{\sum_{k=-K}^{K} \alpha_{\ell^*,k}}{\alpha_{\ell^*,0}}. \tag{3.42}$$

Using the notation α_ℓ^+ and X_t^+ defined respectively in (3.28) and (3.38), we can write

$$\alpha_{\ell^*,0} = \alpha_{\ell^*}^+ \frac{X_{t^*}}{X_{t^*}^+}. \tag{3.43}$$

Substituting (3.43) into (3.40) we obtain

$$\frac{\alpha_{\ell^*,k}}{\alpha_{\ell^*}^+} = \frac{X_{t^*+k}}{X_{t^*}^+}, \qquad -K \le k \le K, \tag{3.44}$$

or equivalently,

$$\alpha_{\ell^*,k} = \alpha_{\ell^*}^+ \frac{X_{t^*+k}}{X_{t^*}^+}, \qquad -K \le k \le K. \tag{3.45}$$

The idea proposed by Smith (2003) for estimating the parameters is to use a clustering algorithm to identify the L signature patterns. Assuming the data are already in the unit Fréchet scale there are four main steps:

1. For each local maximum calculate a candidate signature pattern according to (3.44).
2. Use a clustering algorithm to group all the candidate signature patterns into L clusters.
3. Estimate α_ℓ^+ as the proportion of candidate signature patterns in the ℓth cluster.

4. Obtain estimates of the individual $\alpha_{\ell k}$'s from each cluster center.

Smith (2003) suggested using the K-means clustering algorithm to classify each of the candidate signature pattern into L clusters. The K-means clustering algorithm is described in detail by Hartigan (1975). We use the `kmeans` function implemented in R (R Development Core Team, 2004).

The K-means algorithm provides both the cluster sizes, and the cluster centers that we use in our estimation procedure. Let n_ℓ denote the ℓth cluster size, and denote the cluster centers by the vector $\bar{\boldsymbol{a}}_\ell = (\bar{a}_{\ell,-K}, \ldots, \bar{a}_{\ell,K})$, where

$$\bar{a}_{\ell k} = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} a_k^{(t_i)}, \qquad -K \le k \le K. \tag{3.46}$$

The cluster centers give estimates of the ratios (3.44), and a natural estimator for α_ℓ^+ is $\frac{n_\ell}{n}$. Finally, estimates of the individual parameters (3.45) are obtained by multiplying each cluster center by the corresponding $\frac{n_\ell}{n}$. The details of this method are summarized in Algorithm 3.6.4.

Algorithm 3.6.4 (Clustering-Based Estimation of M3 Processes).

1. Consider a moving window of size $2K+1$, and fix a large threshold $u > 0$.
2. For $t = K+1, \ldots, T-K$, if $x_t > \max\{u, x_{t-K:t+K}\}$ then calculate
$$a_k^{(t)} = \frac{x_{t+k}}{x_t^+}, \qquad -K \le k \le K. \tag{3.47}$$
Denote $\boldsymbol{a}^{(t)} = \left(a_{-K}^{(t)}, \ldots, a_K^{(t)}\right)'$.
3. Use the K-means algorithm to classify each $\boldsymbol{a}^{(t)}$ into one of L clusters. Denote the ℓth cluster size by n_ℓ and the ℓth cluster center by $\bar{\boldsymbol{a}}_\ell$.
4. For each $\ell = 1, \ldots, L$, calculate
$$\hat{\alpha}_\ell^+ = \frac{n_\ell}{n}, \tag{3.48}$$
where $n = \sum_\ell n_\ell$.
5. For each $\ell = 1, \ldots, L$, calculate
$$\hat{\alpha}_{\ell k} = \frac{n_\ell}{n} \bar{a}_{\ell k}, \qquad -K \le k \le K. \tag{3.49}$$

3.6.3 Estimation of M4 Processes

In this section we develop a generalization of Algorithm 3.6.4 for multivariate processes. The main differences from the univariate case are the relative frequencies of the signature patterns and the definition of local maxima.

Consider the M4 process defined by (3.17), i.e.,

$$X_{td} = \max_{1\le\ell\le L}\max_{|k|\le K} \alpha_{\ell kd} Z_{\ell,t-k}, \qquad t\in\mathbb{N}, \quad 1\le d\le D,$$

where $\sum_\ell \sum_k \alpha_{\ell kd} = 1$ for each $d = 1,\dots,D$.

Smith and Weissman (1996) pointed out that for M4 processes the relative frequency of the ℓth signature pattern is *proportional to* $\sum_k \max_d \alpha_{\ell kd}$. This is the key feature of the process that is used for identifying the individual parameters.

Now we define local maxima for the multivariate case. Denote $\boldsymbol{x}_t = (x_{t,1},\dots,x_{t,d})$, so in the usual notation $\boldsymbol{x}_{s:t} = (\boldsymbol{x}_s,\dots,\boldsymbol{x}_t)$ for $s<t$.

Definition 3.6.5 (Local Maximum for Multivariate Processes). For a given integer K and a positive threshold u, consider a block of observations $\boldsymbol{x}_{t-K:t+K},\ K<t<T-K$. We say that x_{td} is a *local maximum* if $x_{td} \ge \max\{u, \boldsymbol{x}_{t-K:t+K}\}$.

Remark. Without loss of generality, our definition assumes $\alpha_{\ell^*,0,d^*} = \max_{(k,d)\in B}\alpha_{\ell^*,k,d}$, where $B = \{(k,d) : |k|\le K,\ 1\le d\le D\}$.

Suppose Z_{ℓ^*,t^*} is much larger than its neighbors, so there is a single index (ℓ^*,t^*), such that

$$X_{t^*+k,d} = \alpha_{\ell^*,k,d} Z_{\ell^*,t^*}, \qquad (k,d)\in B. \tag{3.50}$$

In this case the ℓ^*th signature pattern is a $2K+1$ by D matrix with elements

$$S_{\ell^*,t^*+k,d} = \frac{X_{t^*+k,d}}{X_{t^*,d^*}} = \frac{\alpha_{\ell^*,k,d}}{\alpha_{\ell^*,0,d^*}}, \qquad (k,d)\in B. \tag{3.51}$$

Define the notation

$$X_t^{\ddagger} = \sum_{k=-K}^{K} \max_d X_{t+k,d}, \tag{3.52}$$

$$\alpha_\ell^\ddagger = \sum_{k=-K}^{K} \max_d \alpha_{\ell,k,d}. \tag{3.53}$$

Now note that

$$\sum_{k=-K}^{K} \max_d S_{\ell^*,t^*+k,d} = \frac{X_{t^*}^\ddagger}{X_{t^*,d^*}} = \frac{\alpha_{\ell^*}^\ddagger}{\alpha_{\ell^*,0,d^*}}. \tag{3.54}$$

Therefore

$$\alpha_{\ell^*,0,d^*} = \alpha_{\ell^*}^\ddagger \frac{X_{t^*,d^*}}{X_{t^*}^\ddagger} \tag{3.55}$$

Substituting (3.55) into (3.51) we obtain

$$\frac{\alpha_{\ell^*,k,d}}{\alpha_{\ell^*}^\ddagger} = \frac{X_{t^*+k,d^*}}{X_{t^*}^\ddagger}, \qquad (k,d) \in B, \tag{3.56}$$

or equivalently,

$$\alpha_{\ell^*,k,d} = \alpha_{\ell^*}^\ddagger \frac{X_{t^*+k,d^*}}{X_{t^*}^\ddagger}, \qquad (k,d) \in B. \tag{3.57}$$

In analogy to Algorithm 3.6.4, the estimation procedure for M4 processes is based on calculating (3.56) whenever a local maximum is observed. Then apply the K-means algorithm to obtain both the cluster sizes and the cluster centers. The cluster centers give an estimate of (3.56), and a natural estimator of $\alpha_\ell^\ddagger$ (up to proportionality) is $\frac{n_\ell}{n}$. Estimates of the individual parameters (up to proportionality) are obtained by multiplying the cluster centers by $\frac{n_\ell}{n}$. The details of this method are summarized in Algorithm 3.6.6.

Algorithm 3.6.6 (Clustering-Based Estimation of M4 Processes).

1. Consider a moving window of size $2K+1$, and fix a large threshold $u > 0$.

2. For $t = K+1, \ldots, T-K$, if $x_{td} > \max\{u, \boldsymbol{x}_{t-K:t+K}\}$ then calculate

$$a_{kd}^{(t)} = \frac{x_{t+k,d}}{x_t^{\ddagger}}, \qquad (k,d) \in B. \tag{3.58}$$

   Denote $A^{(t)}$ the matrix with columns $\boldsymbol{a}_d^{(t)} = \left(a_{-K,d}^{(t)}, \ldots, a_{K,d}^{(t)}\right)'$.

3. Use the K-means algorithm to classify each $A^{(t)}$ into one of L clusters. Denote the ℓth cluster size by n_ℓ; and for each $d = 1, \ldots, D$ denote the ℓth cluster center by $\bar{a}_{\ell d}$.

4. For each $\ell = 1, \ldots, L$, calculate

$$\hat{\alpha}_\ell^{\ddagger} \propto \frac{n_\ell}{n}, \tag{3.59}$$

   where $n = \sum_\ell n_\ell$.

5. For each $\ell = 1, \ldots, L$, calculate

$$\hat{\alpha}_{\ell kd} = \frac{n_\ell}{n} \bar{a}_{\ell kd}. \qquad (k,d) \in B. \tag{3.60}$$

6. For each $d = 1, \ldots, D$, normalize so that $\sum_\ell \sum_k \hat{\alpha}_{\ell kd} = 1$.

A justification for the clustering-based approach developed in this and preceding sections is based on Theorem 4.1 of Smith and Weissman (1996). This result essentially states (with changes in notation) that if $x_{t^*,d}$ is a local maximum, then as $u \to \infty$ the probability that (3.50) holds for some ℓ^* tends to 1. The implication of this result is that for sufficiently high u, we can assume that *every* exceedance of u is part of a signature pattern. The clustering method then determines to which one of the L possible patterns a candidate belongs to. The justification of the methods for MM and M3 processes are special cases of this argument.

3.6.4 Simulation Results

We illustrate the clustering-based methods with a simulated path of $T = 4000$ observations of an M3 process with $k = -1, 0, 1$, $L = 3$ signature patterns, and

$$\alpha = \begin{pmatrix} 0.20 & 0.30 & 0.10 \\ 0.05 & 0.15 & 0.10 \\ 0.03 & 0.04 & 0.03 \end{pmatrix}. \tag{3.61}$$

The relative frequency of each signature pattern is given by the sum of the entries in each row of (3.61), i.e., $\sum_k \alpha_{1,k} = 0.6$, $\sum_k \alpha_{2,k} = 0.3$, $\sum_k \alpha_{3,k} = 0.1$. We add normally distributed noise with mean zero and standard deviation σ to the simulated data because in practice the exact signature patterns are never observed.

We estimate the M3 parameters for two levels of noise using a threshold at the 97.5% of the data in both cases. The estimates corresponding to $\sigma = 0.02$ and $\sigma = 0.1$ are denoted $\hat{\alpha}^{(0.02)}$ and $\hat{\alpha}^{(0.1)}$, respectively.

$$\hat{\alpha}^{(0.02)} = \begin{pmatrix} 0.2262 & 0.3416 & 0.1130 \\ 0.0352 & 0.1069 & 0.0707 \\ 0.0319 & 0.0426 & 0.0319 \end{pmatrix}, \qquad \hat{\alpha}^{(0.1)} = \begin{pmatrix} 0.2232 & 0.3463 & 0.1114 \\ 0.0342 & 0.1086 & 0.0700 \\ 0.0320 & 0.0428 & 0.0316 \end{pmatrix}.$$

Figure 3.3 shows the signature patterns of the simulated M3 processes with two different levels of noise. When the noise is very small the clustering algorithm gives almost exactly the actual parameters. When the noise is larger, the candidate signature patterns are more spread out, but the clustering method is still able to identify the clusters correctly.

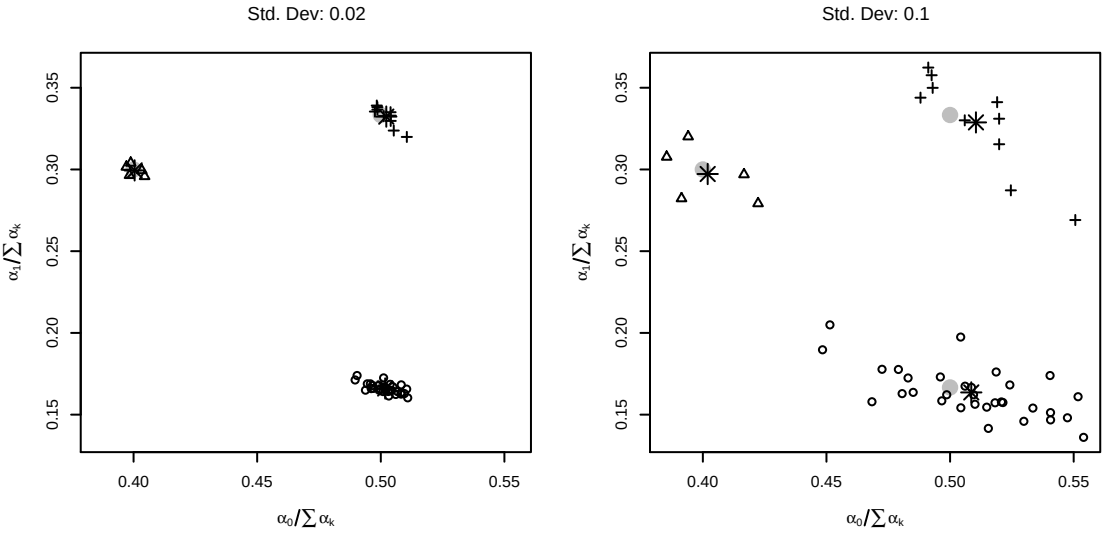

Figure 3.3: *Signature Patterns of Simulated M3 Processes with Noise. The standard deviations of the noise are* $\sigma = 0.02$ *(left) and* $\sigma = 0.1$ *(right). The gray filled circles identify the actual value of* $\alpha_{\ell k} / \sum_k \alpha_{\ell k}$*, and the stars identify the corresponding estimates from the clustering-based method. Candidate signature patterns in the same cluster are identified with empty circles, triangles, and crosses.*

It is important to note that the clustering-based algorithms described in this Chapter require to specify in advance a threshold u, the number of lags K, and the number of signature patterns L. In other words, given u, K, and L, the algorithms classify each of the candidate signature patterns into one of L signature patterns. However these algorithms do not attempt to determine the number of signature patterns, or the number of lags in the model. This question of model selection is a topic for future research.

CHAPTER 4

Markov Property of Moving Maxima Processes

In this Chapter we consider probabilistic properties of two special cases of moving maxima processes, which we refer to as MM(1) and MM(2) processes. Our main results show that MM(1) processes are second-order Markov, and MM(2) processes are third-order Markov. In each case, we provide explicit expressions for the conditional distribution of the process given a sample path up to time t.

In addition, we develop an algorithm to simulate from the conditional distribution of MM(1) processes, and we obtain an explicit expression for the quantile function of this distribution. Applications of these results include sequential Monte Carlo methods to approximate the filtering distributions of MM(1) processes in a state-space model context, as discussed in Chapter 6.

4.1 Conditional Distribution of MM(1) Processes

Consider the MM(1) process

$$X_t = \max\left\{\alpha Z_t,\, (1-\alpha)Z_{t-1}\right\}, \qquad t \in \mathbb{Z}, \tag{4.1}$$

where $\{Z_t\}$ are iid unit Fréchet random variables with distribution function $\Phi_1(x) = e^{-1/x}$, $x > 0$, and $\alpha \in (0, 1)$. We are interested in calculating

$$F_{t+1}(x) \equiv P\left\{X_{t+1} \le x \mid \mathcal{F}_t\right\}, \tag{4.2}$$

where $\mathcal{F}_t \equiv \{X_s = x_s : s \le t\}$. Hence, (4.2) is the conditional distribution of X_{t+1} given a sample path up to time t.

Our main result shows that $F_{t+1}(x)$ may always be calculated as a function of x, x_t and x_{t-1} (but not using values of x_s, $s < t-1$). In other words, MM(1) processes are second-order Markov.

Proposition 4.1.1. *Suppose X_t and $F_{t+1}(x)$ are given by (4.1) and (4.2), respectively.*

If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{1-\alpha}{\alpha} x_t, \\ \exp\left\{-\frac{\alpha}{x}\right\}, & x \geq \frac{1-\alpha}{\alpha} x_t. \end{cases} \tag{4.3}$$

If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then

$$F_{t+1}(x) = \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha}{x_t}\right\}, & 0 \leq x < \frac{1-\alpha}{\alpha} x_t, \\ \exp\left\{-\frac{\alpha}{x}\right\}, & x \geq \frac{1-\alpha}{\alpha} x_t. \end{cases} \tag{4.4}$$

If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then

$$F_{t+1}(x) = \begin{cases} (1-\alpha)\exp\left\{-\frac{1}{x} + \frac{\alpha}{x_t}\right\}, & 0 \leq x < \frac{1-\alpha}{\alpha} x_t, \\ \exp\left\{-\frac{\alpha}{x}\right\}, & x \geq \frac{1-\alpha}{\alpha} x_t. \end{cases} \tag{4.5}$$

Proof. First consider the following four cases:

1. If $X_t = \alpha Z_t > (1-\alpha) Z_{t-1}$ and $X_{t-1} = \alpha Z_{t-1} > (1-\alpha) Z_{t-2}$ then $\frac{X_t}{X_{t-1}} > \frac{1-\alpha}{\alpha}$.
2. If $X_t = (1-\alpha) Z_{t-1} > \alpha Z_t$ and $X_{t-1} = \alpha Z_{t-1} > (1-\alpha) Z_{t-2}$ then $\frac{X_t}{X_{t-1}} = \frac{1-\alpha}{\alpha}$.
3. If $X_t = \alpha Z_t > (1-\alpha) Z_{t-1}$ and $X_{t-1} = (1-\alpha) Z_{t-2} > \alpha Z_{t-1}$, then nothing can be said about the sign of $\frac{X_t}{X_{t-1}} - \frac{1-\alpha}{\alpha}$.
4. If $X_t = (1-\alpha) Z_{t-1} > \alpha Z_t$ and $X_{t-1} = (1-\alpha) Z_{t-2} > \alpha Z_{t-1}$ then $\frac{X_t}{X_{t-1}} < \frac{1-\alpha}{\alpha}$.

From these four cases we deduce:

Case 1: If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then $x_t = \alpha Z_t > (1-\alpha) Z_{t-1}$.

Case 2: If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then $x_t = (1-\alpha) Z_{t-1} > \alpha Z_t$.

Case 3: If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then $x_{t-1} = (1-\alpha)Z_{t-2} > \alpha Z_{t-1}$. However, x_t can be either αZ_t or $(1-\alpha)Z_{t-1}$.

We now consider each of these cases in detail.

Case 1: If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then $Z_t = \frac{x_t}{\alpha}$ and

$$\begin{aligned} P\{X_{t+1} \le x \mid \mathcal{F}_t\} &= P\{\alpha Z_{t+1} \le x,\ (1-\alpha)Z_t \le x \mid \mathcal{F}_t\} \\ &= P\left\{\alpha Z_{t+1} \le x,\ (1-\alpha)Z_t \le x \mid Z_t = \frac{x_t}{\alpha}\right\} \\ &= P\left\{Z_{t+1} \le \frac{\alpha}{x}\right\} \mathbb{I}\left(x \ge \frac{1-\alpha}{\alpha} x_t\right) \\ &= \exp\left\{-\frac{\alpha}{x}\right\} \mathbb{I}\left(x \ge \frac{1-\alpha}{\alpha} x_t\right). \end{aligned} \tag{4.6}$$

Case 2: If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then $Z_t < \frac{x_t}{\alpha}$ and

$$\begin{aligned} P\{X_{t+1} \le x \mid \mathcal{F}_t\} &= P\{\alpha Z_{t+1} \le x,\ (1-\alpha)Z_t \le x \mid \mathcal{F}_t\} \\ &= P\left\{\alpha Z_{t+1} \le x,\ (1-\alpha)Z_t \le x \mid Z_t < \frac{x_t}{\alpha}\right\} \\ &= P\left\{Z_{t+1} \le \frac{\alpha}{x}\right\} P\left\{Z_t \le \frac{x}{1-\alpha} \,\middle|\, Z_t < \frac{x_t}{\alpha}\right\} \\ &= \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha}{x_t}\right\}, & x < \frac{1-\alpha}{\alpha} x_t, \\ \exp\left\{-\frac{\alpha}{x}\right\}, & x \ge \frac{1-\alpha}{\alpha} x_t. \end{cases} \end{aligned} \tag{4.7}$$

Case 3: If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then $Z_{t-1} < \frac{x_{t-1}}{\alpha}$. However, x_t can be either αZ_t or $(1-\alpha)Z_{t-1}$. Then for infinitesimally small δ_t,

$$\begin{aligned} P\left\{Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha},\ Z_{t-1} < \frac{x_t}{1-\alpha} \,\middle|\, \mathcal{F}_t\right\} \\ = P\left\{Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha}, Z_{t-1} < \frac{x_t}{1-\alpha} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha}\right\} \\ = \frac{\alpha\,\delta_t}{x_t^2} \exp\left\{-\frac{1}{x_t} + \frac{\alpha}{x_{t-1}}\right\}, \end{aligned} \tag{4.8}$$

and

$$P\left\{Z_t < \frac{x_t}{\alpha},\ Z_{t-1} \in \frac{(x_t, x_t+\delta_t)}{1-\alpha} \,\middle|\, \mathcal{F}_t\right\}$$

$$= P\left\{ Z_t < \frac{x_t}{\alpha}, Z_{t-1} \in \frac{(x_t, x_t+\delta_t)}{1-\alpha} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha} \right\}$$
$$= \frac{(1-\alpha)\,\delta_t}{x_t^2} \exp\left\{ -\frac{1}{x_t} + \frac{\alpha}{x_{t-1}} \right\}. \tag{4.9}$$

Therefore we have

$$\frac{P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha}, Z_{t-1} < \frac{x_t}{1-\alpha} \,\middle|\, \mathcal{F}_t \right\}}{P\left\{ Z_t < \frac{x_t}{\alpha}, Z_{t-1} \in \frac{(x_t, x_t+\delta_t)}{1-\alpha} \,\middle|\, \mathcal{F}_t \right\}} = \frac{\alpha}{1-\alpha}. \tag{4.10}$$

Since the right-hand side of (4.10) does not depend on x_t and x_{t-1}, it follows that

$$P\{\, \alpha Z_t > (1-\alpha) Z_{t-1} \,|\, \mathcal{F}_t \} = \alpha. \tag{4.11}$$

By combining (4.6), (4.7), and (4.11) we obtain

$$\begin{aligned} P\{\, X_{t+1} \le x \,|\, \mathcal{F}_t\} &= P\left\{ X_{t+1} \le x \,|\, Z_t = \frac{x_t}{\alpha},\ \mathcal{F}_t \right\} P\left\{ Z_t = \frac{x_t}{\alpha} \,\middle|\, \mathcal{F}_t \right\} \\ &\quad + P\left\{ X_{t+1} \le x \,|\, Z_t < \frac{x_t}{\alpha},\ \mathcal{F}_t \right\} P\left\{ Z_t < \frac{x_t}{\alpha} \,\middle|\, \mathcal{F}_t \right\} \\ &= P\left\{ X_{t+1} \le x \,|\, Z_t = \frac{x_t}{\alpha} \right\} \cdot \alpha + P\left\{ X_{t+1} \le x \,|\, Z_t < \frac{x_t}{\alpha} \right\} \cdot (1-\alpha) \\ &= \begin{cases} (1-\alpha) \exp\left\{ -\frac{1}{x} + \frac{\alpha}{x_t} \right\}, & x < \frac{1-\alpha}{\alpha} x_t, \\ \exp\left\{ -\frac{\alpha}{x} \right\}, & x \ge \frac{1-\alpha}{\alpha} x_t. \end{cases} \end{aligned} \tag{4.12}$$

□

Remark. It is clear that X_{t+1} and X_{t-1} are unconditionally independent, because X_{t+1} depends only on Z_t and Z_{t+1}, while X_{t-1} depends only on Z_{t-2} and Z_{t-1}. Similarly, X_{t+1} and X_{t-2} are unconditionally independent. However, conditionally on $\mathcal{F}_t$, X_{t+1} is not independent of X_{t-1}. The surprising result is that conditionally on $\mathcal{F}_t$, X_{t+1} is still independent of X_{t-2}.

The following corollary gives an explicit expression for the conditional probability of a signature pattern (a jump) given $\mathcal{F}_t$.

Corollary 4.1.2. *Suppose X_t and $F_{t+1}(x)$ are given by (4.1) and (4.2), respectively. Let*

$x_t^* = \frac{1-\alpha}{\alpha} x_t$, *then*

$$P\{X_{t+1} = x_t^* \mid \mathcal{F}_t\} = \begin{cases} \exp\left\{-\frac{\alpha}{x_t^*}\right\}, & \frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}, \\ 0, & \frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}, \\ \alpha \exp\left\{-\frac{\alpha}{x_t^*}\right\}, & \frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}. \end{cases} \tag{4.13}$$

Proof. The result follows directly from Proposition 4.1.1 by calculating

$$P\{X_{t+1} = x_t^* \mid \mathcal{F}_t\} = F_{t+1}(x_t^*) - F_{t+1}(x_t^*-), \tag{4.14}$$

where $F_{t+1}(x-) = \lim_{y \uparrow x} F_{t+1}(y) = P\{X_{t+1} < x \mid \mathcal{F}_t\}$. For instance, if $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ we have $F_{t+1}(x_t^*-) = (1-\alpha)\exp\left\{-\frac{\alpha}{x_t^*}\right\}$ and $F_{t+1}(x_t^*) = \exp\left\{-\frac{\alpha}{x_t^*}\right\}$. Therefore, the jump size is $\alpha \exp\left\{-\frac{\alpha}{x_t^*}\right\}$. □

4.1.1 Algorithm for Simulation

In this section we propose an algorithm to simulate from the distribution F_{t+1} derived in Proposition 4.1.1.

First, we introduce the notation used throughout this section. Let Z be a unit Fréchet random variable with distribution function $\Phi_1(x) = e^{-1/x}$, $x > 0$, and let $u > 0$, $\alpha > 0$. The random variable $X = \alpha Z$ is unit Fréchet with scale parameter α. The distribution function of X, denoted $\Phi_{1,\alpha}(x)$, is

$$\Phi_{1,\alpha}(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left\{-\frac{\alpha}{x}\right\}, & x > 0. \end{cases} \tag{4.15}$$

The distribution function of X given that $X < u$ is denoted $\Phi_{1,\alpha}(x \mid x < u)$, where

$$\Phi_{1,\alpha}(x \mid x < u) = \begin{cases} \exp\left\{-\frac{\alpha}{x} + \frac{\alpha}{u}\right\}, & 0 \leq x \leq u, \\ 1, & x > u. \end{cases} \tag{4.16}$$

Similarly, the distribution function of X given that $X > u$ is denoted $\Phi_{1,\alpha}(x \mid x > u)$,

where

$$\Phi_{1,\alpha}\left(x \mid x > u\right) = \begin{cases} 0, & x \leq u, \\ \dfrac{\exp\left\{-\frac{\alpha}{x}\right\} - \exp\left\{-\frac{\alpha}{u}\right\}}{1 - \exp\left\{-\frac{\alpha}{u}\right\}}, & x > u. \end{cases} \tag{4.17}$$

See Appendix A for more details about the properties of Fréchet distributions.

Proposition 4.1.3. *Suppose that $F_{t+1}(x)$ is given as in Proposition 4.1.1. Let $\alpha \in (0,1)$, $x_t > 0$, $x_{t-1} > 0$, and $x_t^* = \frac{1-\alpha}{\alpha} x_t$. The following algorithm generates a random variable $X \sim F_{t+1}(x)$.*

If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then

- *generate $U \sim \mathcal{U}(0,1)$;*
- *if $U \leq e^{-\alpha/x_t^*}$ set $X = x_t^*$;*
- *otherwise, generate $X \sim \Phi_{1,\alpha}\left(x \mid x > x_t^*\right)$.*

If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then

- *generate $U \sim \mathcal{U}(0,1)$;*
- *if $U \leq e^{-\alpha/x_t^*}$ generate $X \sim \Phi_{1,1}\left(x \mid x < x_t^*\right)$;*
- *otherwise, generate $X \sim \Phi_{1,\alpha}\left(x \mid x > x_t^*\right)$.*

If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then

- *generate $U \sim \mathcal{U}(0,1)$;*
- *if $U \leq (1-\alpha)e^{-\alpha/x_t^*}$ generate $X \sim \Phi_{1,\alpha}\left(x \mid x < x_t^*\right)$;*
- *else, if $U \leq e^{-\alpha/x_t^*}$ set $X = x_t^*$;*
- *otherwise, generate $X \sim \Phi_{1,\alpha}\left(x \mid x > x_t^*\right)$.*

Proof. Consider the three cases separately:

Case 1: If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then we need to show that the distribution function of X is (4.3).

Write

$$\begin{aligned} P\left\{X \leq x\right\} &= P\left\{X \leq x \,\middle|\, U \leq e^{-\alpha/x_t^*}\right\} P\left\{U \leq e^{-\alpha/x_t^*}\right\} \\ &\quad + P\left\{X \leq x \,\middle|\, U > e^{-\alpha/x_t^*}\right\} P\left\{U > e^{-\alpha/x_t^*}\right\} \\ &= \mathbb{I}\left(x \geq x_t^*\right) e^{-\alpha/x_t^*} + \Phi_{1,\alpha}\left(x \mid x > x_t^*\right)\left(1 - e^{-\alpha/x_t^*}\right) \end{aligned} \tag{4.18}$$

If $x < x_t^*$ then clearly $P\{X \le x\} = 0$. If $x \ge x_t^*$ then

$$P\{X \le x\} = 1 \cdot e^{-\alpha/x_t^*} + \frac{e^{-\alpha/x} - e^{-\alpha/x_t^*}}{1 - e^{-\alpha/x_t^*}} \left(1 - e^{-\alpha/x_t^*}\right) = e^{-\alpha/x}. \tag{4.19}$$

Case 2: If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then we need to show that the distribution function of X is (4.4). In this case we have

$$P\{X \le x\} = \Phi_{1,1}(x \mid x < x_t^*)\, e^{-\alpha/x_t^*} + \Phi_{1,\alpha}(x \mid x > x_t^*) \left(1 - e^{-\alpha/x_t^*}\right). \tag{4.20}$$

If $x < x_t^*$ then

$$P\{X \le x\} = e^{-1/x+1/x_t^*} e^{-\alpha/x_t^*} + 0 \cdot \left(1 - e^{-\alpha/x_t^*}\right) = e^{-1/x+\alpha/x_t}. \tag{4.21}$$

If $x \ge x_t^*$ then

$$P\{X \le x\} = 1 \cdot e^{-\alpha/x_t^*} + \frac{e^{-\alpha/x} - e^{-\alpha/x_t^*}}{1 - e^{-\alpha/x_t^*}} \left(1 - e^{-\alpha/x_t^*}\right) = e^{-\alpha/x}. \tag{4.22}$$

Case 3: If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then we need to show that the distribution function of X is (4.5). Write

$$\begin{aligned}
P\{X \le x\} &= P\left\{X \le x \,\middle|\, U \le (1-\alpha)e^{-\alpha/x_t^*}\right\} P\left\{U \le (1-\alpha)e^{-\alpha/x_t^*}\right\} \\
&\quad + P\left\{X \le x \,\middle|\, (1-\alpha)e^{-\alpha/x_t^*} < U \le e^{-\alpha/x_t^*}\right\} \\
&\qquad \times P\left\{(1-\alpha)e^{-\alpha/x_t^*} < U \le e^{-\alpha/x_t^*}\right\} \\
&\quad + P\left\{X \le x \,\middle|\, U > e^{-\alpha/x_t^*}\right\} P\left\{U > e^{-\alpha/x_t^*}\right\} \\
&= \Phi_{1,1}(x \mid x < x_t^*)\,(1-\alpha)e^{-\alpha/x_t^*} \\
&\quad + \mathbb{I}(x \ge x_t^*)\,\alpha\, e^{-\alpha/x_t^*} \\
&\quad + \Phi_{1,\alpha}(x \mid x > x_t^*) \left(1 - e^{-\alpha/x_t^*}\right).
\end{aligned} \tag{4.23}$$

If $x < x_t^*$ then

$$\begin{aligned} P\{X \le x\} &= e^{-1/x+1/x_t^*}(1-\alpha)e^{-\alpha/x_t^*} + 0 \cdot \alpha\, e^{-\alpha/x_t^*} + 0 \cdot \left(1 - e^{-\alpha/x_t^*}\right) \\ &= (1-\alpha)e^{-1/x+\alpha/x_t}. \end{aligned} \tag{4.24}$$

If $x \ge x_t^*$ then

$$\begin{aligned} P\{X \le x\} &= 1 \cdot (1-\alpha)e^{-\alpha/x_t^*} + 1 \cdot \alpha\, e^{-\alpha/x_t^*} + \frac{e^{-\alpha/x} - e^{-\alpha/x_t^*}}{1 - e^{-\alpha/x_t^*}} \left(1 - e^{-\alpha/x_t^*}\right) \\ &= e^{-\alpha/x}. \end{aligned} \tag{4.25}$$

□

4.1.2 Quantile Function

Following Embrechts *et al.* (1997, p. 130), suppose h is a non-decreasing function on $\mathbb{R}$. The (left continuous) *generalized inverse of* h is defined as

$$h^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : h(x) \ge p\}. \tag{4.26}$$

By convention, the infimum of an empty set is $+\infty$. The *quantile function* of a distribution function F is

$$F^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : F(x) \ge p\}, \qquad 0 < p < 1. \tag{4.27}$$

The quantity $x_p = F^{\leftarrow}(p)$ defines the p-quantile of F.

Proposition 4.1.4. *Suppose that $F_{t+1}(x)$ is given as in Proposition 4.1.1. Let $\alpha \in (0,1)$, $x_t > 0$, $x_{t-1} > 0$, and $x_t^* = \frac{1-\alpha}{\alpha} x_t$. The quantile function of $F_{t+1}(x)$ is as follows.*

If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} x_t^*, & p \le \exp\left\{-\frac{\alpha}{x_t^*}\right\}, \\ \frac{\alpha}{\log(1/p)}, & p > \exp\left\{-\frac{\alpha}{x_t^*}\right\}. \end{cases} \tag{4.28}$$

If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ *then*

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} \left[\log(1/p) + \frac{\alpha}{x_t}\right]^{-1}, & p \le \exp\left\{-\frac{\alpha}{x_t^*}\right\}, \\ \frac{\alpha}{\log(1/p)}, & p > \exp\left\{-\frac{\alpha}{x_t^*}\right\}. \end{cases} \tag{4.29}$$

If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ *then*

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} \left[\log(1/p) + \frac{\alpha}{x_t} + \log(1-\alpha)\right]^{-1}, & p \le (1-\alpha)\exp\left\{-\frac{\alpha}{x_t^*}\right\}, \\ x_t^*, & (1-\alpha)\exp\left\{-\frac{\alpha}{x_t^*}\right\} < p \le \exp\left\{-\frac{\alpha}{x_t^*}\right\}, \\ \frac{\alpha}{\log(1/p)}, & p > \exp\left\{-\frac{\alpha}{x_t^*}\right\}. \end{cases} \tag{4.30}$$

Proof. Define for $x \ge 0$

$$G_1(x) = \exp\left\{-\frac{\alpha}{x}\right\}, \tag{4.31}$$

$$G_2(x) = \exp\left\{-\frac{1}{x} + \frac{\alpha}{x_t}\right\}, \tag{4.32}$$

$$G_3(x) = \exp\left\{-\frac{1}{x} + \frac{\alpha}{x_t} + \log(1-\alpha)\right\}. \tag{4.33}$$

Then

$$G_1^{\leftarrow}(p) = \frac{\alpha}{\log(1/p)}, \tag{4.34}$$

$$G_2^{\leftarrow}(p) = \left[\log(1/p) + \frac{\alpha}{x_t}\right]^{-1}, \tag{4.35}$$

$$G_3^{\leftarrow}(p) = \left[\log(1/p) + \frac{\alpha}{x_t} + \log(1-\alpha)\right]^{-1}, \tag{4.36}$$

Now consider the three different cases.

Case 1: If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ then $F_{t+1}(x)$ is given by (4.3), so $F_{t+1}(x)$ has a jump at x_t^*. Then

by definition (4.27) we have

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} x_t^*, & p \le F_{t+1}(x_t^*), \\ G_1^{\leftarrow}(p), & p > F_{t+1}(x_t^*). \end{cases} \tag{4.37}$$

Since $F_{t+1}(x_t^*) = \exp\left\{-\frac{\alpha}{x_t^*}\right\}$ we obtain (4.28).

Case 2: If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ then $F_{t+1}(x)$ is given by (4.4), so $F_{t+1}(x)$ is continuous. Thus we have

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} G_2^{\leftarrow}(p), & p \le F_{t+1}(x_t^*), \\ G_1^{\leftarrow}(p), & p > F_{t+1}(x_t^*), \end{cases} \tag{4.38}$$

which gives (4.29).

Case 3: If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ then $F_{t+1}(x)$ is given by (4.5), so $F_{t+1}(x)$ has a jump at x_t^*. Therefore,

$$F_{t+1}^{\leftarrow}(p) = \begin{cases} G_3^{\leftarrow}(p), & p \le F_{t+1}(x_t^*-), \\ x_t^*, & F_{t+1}(x_t^*-) < p \le F_{t+1}(x_t^*), \\ G_1^{\leftarrow}(p), & p > F_{t+1}(x_t^*), \end{cases} \tag{4.39}$$

Since $F_{t+1}(x_t^*-) = (1-\alpha)\exp\left\{-\frac{\alpha}{x_t^*}\right\}$, we obtain (4.30).

□

The median of $F_{t+1}(x)$ is obtained by a straightforward application of Proposition 4.1.4. Observe that

$$\frac{1}{2} \le \exp\left\{-\frac{\alpha}{x_t^*}\right\} \iff x_t^* \ge \frac{\alpha}{\log 2}, \tag{4.40}$$

$$\frac{1}{2} \le (1-\alpha)\exp\left\{-\frac{\alpha}{x_t^*}\right\} \iff x_t^* \ge \frac{\alpha}{\log 2 - \log(1-\alpha)}. \tag{4.41}$$

Corollary 4.1.5. *The median of $F_{t+1}(x)$, denoted ν_{t+1} is as follows.*

If $\frac{x_t}{x_{t-1}} > \frac{1-\alpha}{\alpha}$ *then*

$$\nu_{t+1} = \begin{cases} \frac{\alpha}{\log 2}, & x_t^* < \frac{\alpha}{\log 2}, \\ x_t^*, & x_t^* \geq \frac{\alpha}{\log 2}. \end{cases} \tag{4.42}$$

If $\frac{x_t}{x_{t-1}} = \frac{1-\alpha}{\alpha}$ *then*

$$\nu_{t+1} = \begin{cases} \frac{\alpha}{\log 2}, & x_t^* < \frac{\alpha}{\log 2}, \\ \left[\log 2 + \frac{\alpha}{x_t}\right]^{-1}, & x_t^* \geq \frac{\alpha}{\log 2}. \end{cases} \tag{4.43}$$

If $\frac{x_t}{x_{t-1}} < \frac{1-\alpha}{\alpha}$ *then*

$$\nu_{t+1} = \begin{cases} \frac{\alpha}{\log 2}, & x_t^* < \frac{\alpha}{\log 2}, \\ x_t^*, & \frac{\alpha}{\log 2} \leq x_t^* < \frac{\alpha}{\log 2 - \log(1-\alpha)}, \\ \left[\log 2 + \frac{\alpha}{x_t} + \log(1-\alpha)\right]^{-1}, & x_t^* \geq \frac{\alpha}{\log 2 - \log(1-\alpha)}. \end{cases} \tag{4.44}$$

4.2 Conditional Distribution of MM(2) Processes

Consider the MM(2) process

$$X_t = \max\left\{\alpha_0 Z_t,\, \alpha_1 Z_{t-1},\, \alpha_2 Z_{t-2}\right\}, \qquad t \in \mathbb{Z}, \tag{4.45}$$

where $\{Z_t\}$ are iid unit Fréchet random variables with distribution function $\Phi_1(x) = e^{-1/x}$, $x > 0$, and $\alpha_k \geq 0$ satisfying $\sum_k \alpha_k = 1$.

Our main goal in this section is to show that MM(2) processes are third-order Markov. Following the developments in §4.1, we are interested in calculating

$$F_{t+1}(x) \equiv P\left\{X_{t+1} \leq x \mid \mathcal{F}_t\right\}. \tag{4.46}$$

We find that (4.46) depends on the relationship between α_1/α_0 and α_2/α_1, and also depends on past values of the process. In what follows we provide explicit expressions for (4.46) in each of these cases. We defer the proofs to §4.2.1.

Result for $\alpha_2/\alpha_1 < \alpha_1/\alpha_0$

If $\alpha_2/\alpha_1 < \alpha_1/\alpha_0$, we claim that $F_{t+1}(x)$ may always be calculated as a function of x, x_t, and x_{t-1} (but not using values of x_s, $s < t-1$). In other words, under this restriction MM(2) processes are second-order Markov.

Proposition 4.2.1. *Suppose $\alpha_2/\alpha_1 < \alpha_1/\alpha_0$, and X_t and $F_{t+1}(x)$ are given by (4.45) and (4.46), respectively.*

If $x_t/x_{t-1} > \alpha_1/\alpha_0$ then

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.47}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ then

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{-\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t}\right\}, & \frac{\alpha_2}{\alpha_1} x_t \leq x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.48}$$

If $\alpha_2/\alpha_1 < x_t/x_{t-1} < \alpha_1/\alpha_0$ then

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \frac{\alpha_1}{\alpha_0+\alpha_1} \exp\left\{-\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t}\right\}, & \frac{\alpha_2}{\alpha_1} x_t \leq x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.49}$$

If $x_t/x_{t-1} = \alpha_2/\alpha_1$ then

$$F_{t+1}(x) = \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{-\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t}\right\}, & \frac{\alpha_2}{\alpha_1} x_t \leq x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.50}$$

If $x_t/x_{t-1} < \alpha_2/\alpha_1$ *then*

$$F_{t+1}(x) = \begin{cases} \alpha_2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_2}{\alpha_1}x_t, \\ (\alpha_1+\alpha_2)\exp\left\{-\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t}\right\}, & \frac{\alpha_2}{\alpha_1}x_t \le x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \tag{4.51}$$

Result for $\alpha_2/\alpha_1 > \alpha_1/\alpha_0$

If $\alpha_2/\alpha_1 > \alpha_1/\alpha_0$, we claim that $F_{t+1}(x)$ may always be calculated as a function of x, x_t, x_{t-1}, and x_{t-2} (but not using values of x_s, $s < t-2$). In other words, MM(2) processes are third-order Markov.

Proposition 4.2.2. *Suppose* $\alpha_2/\alpha_1 > \alpha_1/\alpha_0$*, and* X_t *and* $F_{t+1}(x)$ *are given by* (4.45) *and* (4.46)*, respectively.*

If $x_t/x_{t-1} > \alpha_2/\alpha_1$ *then*

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \tag{4.52}$$

If $x_t/x_{t-1} = \alpha_2/\alpha_1$ *then*

$$F_{t+1}(x) = \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha_0}{x_t} + \frac{\alpha_0}{x_{t-1}}\right\}, & x < \frac{\alpha_1}{\alpha_0}x_t = \frac{\alpha_2}{\alpha_0}x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_1}{\alpha_0}x_t = \frac{\alpha_2}{\alpha_0}x_{t-1}. \end{cases} \tag{4.53}$$

If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ *and* $x_{t-1}/x_{t-2} > \alpha_2/\alpha_1$ *then*

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_0}x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_2}{\alpha_0}x_{t-1}. \end{cases} \tag{4.54}$$

If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ *and* $x_{t-1}/x_{t-2} \leq \alpha_2/\alpha_1$ *then*

$$F_{t+1}(x) = \begin{cases} \left(\frac{\alpha_2}{\alpha_0+\alpha_2}\right)^2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0}{x_t} + \frac{\alpha_0}{x_{t-1}}\right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \frac{\alpha_2}{\alpha_0+\alpha_2} \exp\left\{-\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_0}{x_{t-1}}\right\}, & \frac{\alpha_1}{\alpha_0} x_t \leq x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \tag{4.55}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *then*

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t = \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_2}{\alpha_1} x_t = \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \tag{4.56}$$

If $x_t/x_{t-1} < \alpha_1/\alpha_0$ *then*

$$F_{t+1}(x) = \begin{cases} \alpha_2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ (\alpha_0+\alpha_2) \exp\left\{-\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_1}{x_t}\right\}, & \frac{\alpha_1}{\alpha_0} x_t \leq x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_2}{\alpha_1} x_t. \end{cases} \tag{4.57}$$

Result for $\alpha_2/\alpha_1 = \alpha_1/\alpha_0$

If $\alpha_2/\alpha_1 = \alpha_1/\alpha_0$, we claim that $F_{t+1}(x)$ may always be calculated as a function of x, x_t, x_{t-1}, and x_{t-2} (but not using values of $x_s,\ s < t-2$). In other words, MM(2) processes are third-order Markov.

Proposition 4.2.3. *Suppose* $\alpha_2/\alpha_1 = \alpha_1/\alpha_0$*, and* X_t *and* $F_{t+1}(x)$ *are given by* (4.45) *and* (4.46)*, respectively.*

If $x_t/x_{t-1} > \alpha_1/\alpha_0$*; or* $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} > \alpha_1/\alpha_0$*; then*

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.58}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} = \alpha_1/\alpha_0$ *then*

$$F_{t+1}(x) = \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \tag{4.59}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} < \alpha_1/\alpha_0$ *then*

$$F_{t+1}(x) = \begin{cases} \frac{\alpha_1}{\alpha_0+\alpha_1}\exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \tag{4.60}$$

If $x_t/x_{t-1} < \alpha_1/\alpha_0$ *then*

$$F_{t+1}(x) = \begin{cases} \alpha_2\exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \tag{4.61}$$

Conditional Probabilities of Jumps

Corollaries 4.2.4, 4.2.5, and 4.2.6 give explicit expressions for the conditional probabilities of a jump in the process given $\mathcal{F}_t$. We use the notation $x^*_{t,0} = \frac{\alpha_2}{\alpha_0}x_{t-1}$, $x^*_{t,1} = \frac{\alpha_1}{\alpha_0}x_t$, $x^*_{t,2} = \frac{\alpha_2}{\alpha_1}x_t$. These results follow directly by calculating

$$p_i = P\left\{X_{t+1} = x^*_{t,i} \,\middle|\, \mathcal{F}_t\right\} = F_{t+1}(x^*_{t,i}) - F_{t+1}(x^*_{t,i}-), \qquad i = 0, 1, 2. \tag{4.62}$$

Note that for $\alpha_2/\alpha_1 \leq \alpha_1/\alpha_0$ we have $p_0 = 0$.

Corollary 4.2.4. *Suppose* $\alpha_2/\alpha_1 < \alpha_1/\alpha_0$*, and* X_t *and* $F_{t+1}(x)$ *are given by* (4.45) *and* (4.46)*, respectively.*

If $x_t/x_{t-1} > \alpha_1/\alpha_0$ *then* $p_2 = 0$*, and*

$$p_1 = \exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}. \tag{4.63}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *then* $p_1 = 0$, *and*

$$p_2 = \exp\left\{-\frac{(\alpha_0+\alpha_1)^2 - \alpha_0}{\alpha_1\, x^*_{t,2}}\right\}. \tag{4.64}$$

If $\alpha_2/\alpha_1 < x_t/x_{t-1} < \alpha_1/\alpha_0$ *then*

$$p_1 = \frac{\alpha_0}{\alpha_0+\alpha_1}\exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}, \tag{4.65}$$

$$p_2 = \frac{\alpha_1}{\alpha_0+\alpha_1}\exp\left\{-\frac{(\alpha_0+\alpha_1)^2 - \alpha_0}{\alpha_1\, x^*_{t,2}}\right\}. \tag{4.66}$$

If $x_t/x_{t-1} = \alpha_2/\alpha_1$ *then* $p_1 = p_2 = 0$.

If $x_t/x_{t-1} < \alpha_2/\alpha_1$ *then*

$$p_1 = \alpha_0\, \exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}, \tag{4.67}$$

$$p_2 = \alpha_1\, \exp\left\{-\frac{(\alpha_0+\alpha_1)^2 - \alpha_0}{\alpha_1\, x^*_{t,2}}\right\}. \tag{4.68}$$

Corollary 4.2.5. *Suppose* $\alpha_2/\alpha_1 > \alpha_1/\alpha_0$, *and* X_t *and* $F_{t+1}(x)$ *are given by* (4.45) *and* (4.46), *respectively.*

If $x_t/x_{t-1} > \alpha_2/\alpha_1$ *then* $p_0 = p_2 = 0$, *and*

$$p_1 = \exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}. \tag{4.69}$$

If $x_t/x_{t-1} = \alpha_2/\alpha_1$ *then* $p_0 = p_1 = p_2 = 0$.

If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ *and* $x_{t-1}/x_{t-2} > \alpha_2/\alpha_1$ *then* $p_1 = p_2 = 0$, *and*

$$p_0 = \exp\left\{-\frac{\alpha_0}{x^*_{t,0}}\right\}. \tag{4.70}$$

If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ *and* $x_{t-1}/x_{t-2} \le \alpha_2/\alpha_1$ *then* $p_2 = 0$, *and*

$$p_0 = \frac{\alpha_0}{\alpha_0+\alpha_2}\exp\left\{-\frac{\alpha_0}{x^*_{t,0}}\right\}, \tag{4.71}$$

$$p_1 = \frac{\alpha_0\,\alpha_2}{(\alpha_0+\alpha_2)^2}\exp\left\{-\frac{\alpha_0+\alpha_2}{x^*_{t,1}}+\frac{\alpha_0}{x_{t-1}}\right\}. \tag{4.72}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *then* $p_1 = 0$, *and*

$$p_0 = p_2 = \exp\left\{-\frac{\alpha_0}{x^*_{t,2}}\right\}. \tag{4.73}$$

If $x_t/x_{t-1} < \alpha_1/\alpha_0$ *then* $p_0 = 0$, *and*

$$p_1 = \alpha_0\,\exp\left\{-\frac{\alpha_0+\alpha_2}{x^*_{t,1}}+\frac{\alpha_2}{x^*_{t,2}}\right\}, \tag{4.74}$$

$$p_2 = \alpha_1\,\exp\left\{-\frac{\alpha_0}{x^*_{t,2}}\right\}. \tag{4.75}$$

Corollary 4.2.6. *Suppose* $\alpha_2/\alpha_1 = \alpha_1/\alpha_0$, *and* X_t *and* $F_{t+1}(x)$ *are given by* (4.45) *and* (4.46), *respectively.*

If $x_t/x_{t-1} > \alpha_1/\alpha_0$; *or* $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} > \alpha_1/\alpha_0$; *then*

$$p_1 = p_2 = \exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}. \tag{4.76}$$

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} = \alpha_1/\alpha_0$ *then* $p_1 = p_2 = 0$.

If $x_t/x_{t-1} = \alpha_1/\alpha_0$ *and* $x_{t-1}/x_{t-2} < \alpha_1/\alpha_0$ *then*

$$p_1 = p_2 = \frac{\alpha_0}{\alpha_0+\alpha_1}\exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}. \tag{4.77}$$

If $x_t/x_{t-1} < \alpha_1/\alpha_0$ *then*

$$p_1 = p_2 = (\alpha_0+\alpha_1)\exp\left\{-\frac{\alpha_0}{x^*_{t,1}}\right\}. \tag{4.78}$$

4.2.1 Proofs

Proof of Proposition 4.2.1.

Assume $\alpha_2/\alpha_1 < \alpha_1/\alpha_0$. Consider all possible scenarios for the ratio X_t/X_{t-1} as shown in Table 4.1. We deduce that the ratio X_t/X_{t-1} can fall into one of the five intervals defined

by the values $(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$, and Table 4.2 summarizes these five cases. In what follows we consider each of these cases in detail.

X_t	X_{t-1}	X_t/X_{t-1}
$\alpha_0 Z_t$	$\alpha_0 Z_{t-1}$	$> \alpha_1/\alpha_0$
	$\alpha_1 Z_{t-2}$	$> \alpha_2/\alpha_1$
	$\alpha_2 Z_{t-3}$	?
$\alpha_1 Z_{t-1}$	$\alpha_0 Z_{t-1}$	$= \alpha_1/\alpha_0$
	$\alpha_1 Z_{t-2}$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$
	$\alpha_2 Z_{t-3}$	$< \alpha_1/\alpha_0$
$\alpha_2 Z_{t-2}$	$\alpha_0 Z_{t-1}$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$
	$\alpha_1 Z_{t-2}$	$= \alpha_2/\alpha_1$
	$\alpha_2 Z_{t-3}$	$< \alpha_2/\alpha_1$

Table 4.1: *Ratio of Contiguous Observations from MM(2) Processes for the Proof of Proposition 4.2.1.*

Case	x_t/x_{t-1}	Deduction
1	$> \alpha_1/\alpha_0$	$Z_t = \frac{x_t}{\alpha_0}$ and $Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0}$
2	$= \alpha_1/\alpha_0$	$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$
3	$\in (\alpha_2/\alpha_1, \alpha_1/\alpha_0)$	$Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_1 Z_{t-1}$.
4	$= \alpha_2/\alpha_1$	$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} < \frac{x_t}{\alpha_1}$
5	$< \alpha_2/\alpha_1$	$Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$.

Table 4.2: *Deduced Cases for the Proof of Proposition 4.2.1.*

Case 1: If $x_t/x_{t-1} > \alpha_1/\alpha_0$, then $Z_t = \frac{x_t}{\alpha_0}$ and $Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$. Note that $\frac{x_t}{x_{t-1}} > \frac{\alpha_1}{\alpha_0} > \frac{\alpha_2}{\alpha_1} \iff \frac{\alpha_1}{\alpha_0} x_t > \frac{\alpha_2}{\alpha_0} x_{t-1}$, and $P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} = 1$ for $x \geq \frac{\alpha_1}{\alpha_0} x_t$. Then

$$
\begin{aligned}
F_{t+1}(x) &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0},\ Z_t \leq \frac{x}{\alpha_1},\ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} \\
&= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0} x_t \right) P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} \\
&= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0} x_t \right)
\end{aligned}
$$

$$= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.79}$$

Case 2: If $x_t/x_{t-1} = \alpha_1/\alpha_0$, then $Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$. We also have $\frac{\alpha_1}{\alpha_0} x_t > \frac{\alpha_2}{\alpha_1} x_t = \frac{\alpha_2}{\alpha_0} x_{t-1}$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0},\ Z_t \leq \frac{x}{\alpha_1},\ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_t}{\alpha_1} \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} P\left\{ Z_t \leq \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_2}{\alpha_1} x_t \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t} \right\}, & \frac{\alpha_2}{\alpha_1} x_t \leq x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \end{aligned} \tag{4.80}$$

Case 3: If $\alpha_2/\alpha_1 < x_t/x_{t-1} < \alpha_1/\alpha_0$, then $Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$ and $Z_{t-2} \leq \frac{x_{t-1}}{\alpha_1} = \min\left(\frac{x_{t-1}}{\alpha_1}, \frac{x_t}{\alpha_2} \right)$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_1 Z_{t-1}$. Then for infinitesimally small δ_t,

$$\begin{aligned} &P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, \mathcal{F}_t \right\} \\ &\quad = P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0},\ Z_{t-2} \leq \frac{x_{t-1}}{\alpha_1} \right\} \\ &\quad = P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0} \right\} P\left\{ Z_{t-1} < \frac{x_t}{\alpha_1} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0} \right\} \\ &\quad = \frac{\alpha_0 \delta_t}{x_t^2} \exp\left\{ -\frac{\alpha_0+\alpha_1}{x_t} + \frac{\alpha_0}{x_{t-1}} \right\}. \end{aligned} \tag{4.81}$$

Similarly,

$$\begin{aligned} &P\left\{ Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} \in \frac{(x_t, x_t+\delta_t)}{\alpha_1},\ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, \mathcal{F}_t \right\} \\ &\qquad\qquad = \frac{\alpha_1 \delta_t}{x_t^2} \exp\left\{ -\frac{\alpha_0+\alpha_1}{x_t} + \frac{\alpha_0}{x_{t-1}} \right\}. \end{aligned} \tag{4.82}$$

Define $A = \{\alpha_0 Z_t > \alpha_1 Z_{t-1}\}$. Since $Z_{t-2} < \frac{x_t}{\alpha_2}$ then $\{X_t = x_t\} = A \cup A^c$, and we

have

$$\begin{aligned} P\{A \mid \mathcal{F}_t\} &= P\{A \mid X_t = x_t,\ \mathcal{F}_t\} \\ &= \frac{P\{A \mid \mathcal{F}_t\}}{P\{A \cup A^c \mid \mathcal{F}_t\}} \\ &= \frac{\alpha_0}{\alpha_0 + \alpha_1}. \end{aligned} \tag{4.83}$$

Thus we can write

$$F_{t+1}(x) = \frac{\alpha_0}{\alpha_0 + \alpha_1} F_{t+1}^{(0)}(x) + \frac{\alpha_1}{\alpha_0 + \alpha_1} F_{t+1}^{(1)}(x), \tag{4.84}$$

where

$$F_{t+1}^{(i)}(x) = P\left\{ X_{t+1} \le x \;\middle|\; Z_{t-i} = \frac{x_t}{\alpha_i},\ \mathcal{F}_t \right\}, \qquad i = 0, 1, 2. \tag{4.85}$$

Note that If $Z_t = \frac{x_t}{\alpha_0}$ then $Z_{t-1} < \frac{x_t}{\alpha_1} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$ (because $\frac{x_t}{x_{t-1}} < \frac{\alpha_1}{\alpha_0}$), so that $F_{t+1}^{(0)}(x)$ is given by (4.79). On the other hand, if $Z_{t-1} = \frac{x_t}{\alpha_1}$ then $F_{t+1}^{(1)}(x)$ is given by (4.80). By combining these results into (4.84) we obtain

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \frac{\alpha_1}{\alpha_0 + \alpha_1} \exp\left\{ -\frac{\alpha_0 + \alpha_1}{x} + \frac{\alpha_0}{x_t} \right\}, & \frac{\alpha_2}{\alpha_1} x_t \le x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.86}$$

Case 4: If $x_t / x_{t-1} = \alpha_2 / \alpha_1$ then $Z_{t-2} = \frac{x_t}{\alpha_2} = \frac{x_{t-1}}{\alpha_1}$, $Z_t < \frac{x_t}{\alpha_0}$, and since $\frac{x_t}{x_{t-1}} = \frac{\alpha_2}{\alpha_1} < \frac{\alpha_1}{\alpha_0} \iff \frac{x_t}{\alpha_1} < \frac{x_{t-1}}{\alpha_0}$, then $Z_{t-1} < \frac{x_t}{\alpha_1} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$. Write

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0},\ Z_t \le \frac{x}{\alpha_1},\ Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1} \right\} \\ &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0} \right\} P\left\{ Z_t \le \frac{x}{\alpha_1} \;\middle|\; Z_t < \frac{x_t}{\alpha_0} \right\} P\left\{ Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_{t-1} < \frac{x_t}{\alpha_1} \right\}. \end{aligned} \tag{4.87}$$

The second product term in (4.87) is

$$P\left\{ Z_t \le \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} = \begin{cases} \exp\left\{ -\frac{\alpha_1}{x} + \frac{\alpha_0}{x_t} \right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ 1, & x \ge \frac{\alpha_1}{\alpha_0} x_t, \end{cases} \tag{4.88}$$

and the third product term in (4.87) is

$$P\left\{ Z_{t-1} \le \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_t}{\alpha_1} \right\} = \begin{cases} \exp\left\{ -\frac{\alpha_2}{x} + \frac{\alpha_1}{x_t} \right\}, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ 1, & x \ge \frac{\alpha_2}{\alpha_1} x_t. \end{cases} \tag{4.89}$$

Therefore

$$F_{t+1}(x) = \begin{cases} \exp\left\{ -\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t} \right\}, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t} \right\}, & \frac{\alpha_2}{\alpha_1} x_t \le x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.90}$$

Case 5: If $x_t/x_{t-1} < \alpha_2/\alpha_1$ then $Z_{t-3} = \frac{x_{t-1}}{\alpha_2}$ and $x_{t-1}/x_{t-2} \le \alpha_2/\alpha_1 < \alpha_1/\alpha_0$. Hence $Z_{t-2} < \frac{x_{t-1}}{\alpha_1} = \min\left(\frac{x_{t-1}}{\alpha_1}, \frac{x_{t-2}}{\alpha_0} \right)$ and $Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$. Then for infinitesimally small δ_t,

$$\begin{aligned} P\Bigg\{ Z_t \in & \frac{(x_t, x_t+\delta_t)}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ Z_{t-2} < \frac{x_t}{\alpha_2} \,\Bigg|\, \mathcal{F}_t \Bigg\} \\ &= P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0} \right\} P\left\{ Z_{t-1} < \frac{x_t}{\alpha_1} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0} \right\} \\ &\quad \times P\left\{ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, Z_{t-2} < \frac{x_{t-1}}{\alpha_1} \right\} \\ &= \frac{\alpha_0 \delta_t}{x_t^2} \exp\left\{ -\frac{1}{x_t} + \frac{\alpha_0+\alpha_1}{x_{t-1}} \right\}. \end{aligned} \tag{4.91}$$

Similar calculations show

$$P\left\{ Z_{t-i} \in \frac{(x_t, x_t+\delta_t)}{\alpha_i},\ Z_{t-j} < \frac{x_t}{\alpha_j},\ j \ne i \,\middle|\, \mathcal{F}_t \right\} = \frac{\alpha_i \delta_t}{x_t^2} \exp\left\{ -\frac{1}{x_t} + \frac{\alpha_0+\alpha_1}{x_{t-1}} \right\}, \quad i = 1, 2. \tag{4.92}$$

Then it follows that

$$P\left\{\alpha_i Z_{t-i} > \alpha_k Z_{t-k},\ k \neq i \,\middle|\, Z_{t-3} = \frac{x_{t-1}}{\alpha_2},\ \mathcal{F}_t\right\} = \alpha_i, \qquad i = 0, 1, 2. \tag{4.93}$$

Thus we can write

$$F_{t+1}(x) = \alpha_0 F_{t+1}^{(0)}(x) + \alpha_1 F_{t+1}^{(1)}(x) + \alpha_2 F_{t+1}^{(2)}(x), \tag{4.94}$$

where $F_{t+1}^{(i)}(x)$ is defined by (4.85).

On the other hand, $F_{t+1}^{(0)}(x)$ is given by (4.79), $F_{t+1}^{(1)}(x)$ is given by (4.80), and $F_{t+1}^{(2)}(x)$ is given by (4.90). Substituting into (4.94) gives

$$F_{t+1}(x) = \begin{cases} \alpha_2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ (\alpha_1+\alpha_2) \exp\left\{-\frac{\alpha_0+\alpha_1}{x} + \frac{\alpha_0}{x_t}\right\}, & \frac{\alpha_2}{\alpha_1} x_t \leq x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.95}$$

□

Proof of Proposition 4.2.2.

Assume $\alpha_2/\alpha_1 > \alpha_1/\alpha_0$. Consider all possible scenarios for the ratios X_t/X_{t-1} and X_{t-1}/X_{t-2} as shown in Table 4.3. We deduce six cases as summarized in Table 4.4, and consider each of these cases in detail.

X_t	X_{t-1}	X_{t-2}	X_t/X_{t-1}	X_{t-1}/X_{t-2}
$\alpha_0 Z_t$	$\alpha_0 Z_{t-1}$	$\alpha_0 Z_{t-2}$	$> \alpha_1/\alpha_0$	$> \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$> \alpha_1/\alpha_0$	$> \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	$> \alpha_1/\alpha_0$	?
	$\alpha_1 Z_{t-2}$	$\alpha_0 Z_{t-2}$	$> \alpha_2/\alpha_1$	$= \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$> \alpha_2/\alpha_1$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$
		$\alpha_2 Z_{t-4}$	$> \alpha_2/\alpha_1$	$< \alpha_1/\alpha_0$
	$\alpha_2 Z_{t-3}$	$\alpha_0 Z_{t-2}$	?	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$
		$\alpha_1 Z_{t-3}$	?	$= \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	?	$< \alpha_2/\alpha_1$

Table 4.3: *Ratio of Contiguous Observations from MM(2) Processes for the Proof of Propositions 4.2.2 and 4.2.3.*

X_t	X_{t-1}	X_{t-2}	X_t/X_{t-1}	X_{t-1}/X_{t-2}
$\alpha_1 Z_{t-1}$	$\alpha_0 Z_{t-1}$	$\alpha_0 Z_{t-2}$	$= \alpha_1/\alpha_0$	$> \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$= \alpha_1/\alpha_0$	$> \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	$= \alpha_1/\alpha_0$	?
	$\alpha_1 Z_{t-2}$	$\alpha_0 Z_{t-2}$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$	$= \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$
		$\alpha_2 Z_{t-4}$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$	$< \alpha_1/\alpha_0$
	$\alpha_2 Z_{t-3}$	$\alpha_0 Z_{t-2}$	$< \alpha_1/\alpha_0$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$
		$\alpha_1 Z_{t-3}$	$< \alpha_1/\alpha_0$	$= \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	$< \alpha_1/\alpha_0$	$< \alpha_2/\alpha_1$
$\alpha_2 Z_{t-2}$	$\alpha_0 Z_{t-1}$	$\alpha_0 Z_{t-2}$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$	$> \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$	$> \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$	?
	$\alpha_1 Z_{t-2}$	$\alpha_0 Z_{t-2}$	$= \alpha_2/\alpha_1$	$= \alpha_1/\alpha_0$
		$\alpha_1 Z_{t-3}$	$= \alpha_2/\alpha_1$	$(\alpha_2/\alpha_1, \alpha_1/\alpha_0)$
		$\alpha_2 Z_{t-4}$	$= \alpha_2/\alpha_1$	$< \alpha_1/\alpha_0$
	$\alpha_2 Z_{t-3}$	$\alpha_0 Z_{t-2}$	$< \alpha_2/\alpha_1$	$(\alpha_1/\alpha_0, \alpha_2/\alpha_1)$
		$\alpha_1 Z_{t-3}$	$< \alpha_2/\alpha_1$	$= \alpha_2/\alpha_1$
		$\alpha_2 Z_{t-4}$	$< \alpha_2/\alpha_1$	$< \alpha_2/\alpha_1$

Table 4.3: *(continued) Ratio of Contiguous Observations from MM(2) Processes for the Proof of Propositions 4.2.2 and 4.2.3.*

Case	x_t/x_{t-1}	x_{t-1}/x_{t-2}	Deduction
1	$> \alpha_2/\alpha_1$		$Z_t = \frac{x_t}{\alpha_0}$ and $Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0}$.
2	$= \alpha_2/\alpha_1$		$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$.
3.1	$\in (\alpha_1/\alpha_0, \alpha_2/\alpha_1)$	$> \alpha_2/\alpha_1$	$Z_{t-1} = \frac{x_{t-1}}{\alpha_0} < \frac{x_t}{\alpha_1}$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_2 Z_{t-2}$.
3.2	$\in (\alpha_1/\alpha_0, \alpha_2/\alpha_1)$	$\leq \alpha_2/\alpha_1$	$Z_{t-1} < \frac{x_t}{\alpha_1}$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_2 Z_{t-2}$, and X_{t-1} can be either $\alpha_0 Z_{t-1}$ or $\alpha_2 Z_{t-3}$.
4	$= \alpha_1/\alpha_0$		$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$.
5	$< \alpha_1/\alpha_0$		$Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$.

Table 4.4: *Deduced Cases for the Proof of Proposition 4.2.2.*

Case 1: If $x_t/x_{t-1} > \alpha_2/\alpha_1$ then $Z_t = \frac{x_t}{\alpha_0}$. Since $\frac{x_t}{x_{t-1}} > \frac{\alpha_2}{\alpha_1} > \frac{\alpha_1}{\alpha_0}$ then $Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} =$

$\min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$ and $\frac{\alpha_1}{\alpha_0}x_t > \frac{\alpha_2}{\alpha_0}x_{t-1}$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{X_{t+1} \le x \;\middle|\; Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} \le \frac{x_{t-1}}{\alpha_0}\right\} \\ &= P\left\{Z_{t+1} \le \frac{x}{\alpha_0}\right\} \mathbb{I}\left(x \ge \frac{\alpha_1}{\alpha_0}x_t\right) P\left\{Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_{t-1} \le \frac{x_{t-1}}{\alpha_0}\right\} \\ &= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \end{aligned} \tag{4.96}$$

Case 2: If $x_t/x_{t-1} = \alpha_2/\alpha_1$ then $Z_{t-2} = \frac{x_t}{\alpha_2} = \frac{x_{t-1}}{\alpha_1}$, so $Z_t < \frac{x_t}{\alpha_0}$. Moreover, $\frac{x_t}{x_{t-1}} = \frac{\alpha_2}{\alpha_1} > \frac{\alpha_1}{\alpha_0}$ implies $Z_{t-1} < \frac{x_{t-1}}{\alpha_0} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{Z_{t+1} \le \frac{x}{\alpha_0},\ Z_t \le \frac{x}{\alpha_1},\ Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_{t-1}}{\alpha_0}\right\} \\ &= P\left\{Z_{t+1} \le \frac{x}{\alpha_0}\right\} P\left\{Z_t \le \frac{x}{\alpha_1} \;\middle|\; Z_t < \frac{x_t}{\alpha_0}\right\} P\left\{Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_{t-1} < \frac{x_{t-1}}{\alpha_0}\right\} \\ &= \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha_0}{x_t} + \frac{\alpha_0}{x_{t-1}}\right\}, & x < \frac{\alpha_1}{\alpha_0}x_t = \frac{\alpha_2}{\alpha_0}x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_1}{\alpha_0}x_t = \frac{\alpha_2}{\alpha_0}x_{t-1}. \end{cases} \end{aligned} \tag{4.97}$$

Case 3.1: If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ and $x_{t-1}/x_{t-2} > \alpha_2/\alpha_1$ then $Z_{t-1} = \frac{x_{t-1}}{\alpha_0} < \frac{x_t}{\alpha_1}$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_2 Z_{t-2}$. Also note that $\frac{\alpha_1}{\alpha_0}x_t < \frac{\alpha_2}{\alpha_0}x_{t-1}$.

We introduce the following notation.

$$F_{t+1}^{(i,j)}(x) = P\left\{X_{t+1} \le x \;\middle|\; Z_{t-i} = \frac{x_t}{\alpha_i},\ Z_{t-j} = \frac{x_{t-1}}{\alpha_j},\ \mathcal{F}_t\right\}, \qquad i = 0,1,2,\ j = 0,1,2. \tag{4.98}$$

If $Z_t = \frac{x_t}{\alpha_0}$ then

$$\begin{aligned} F_{t+1}^{(0,0)}(x) &= P\left\{Z_{t+1} \le \frac{x}{\alpha_0},\ Z_t \le \frac{x}{\alpha_1},\ Z_{t-1} \le \frac{x}{\alpha_2} \;\middle|\; Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_{t-1}}{\alpha_0},\ \mathcal{F}_t\right\} \\ &= P\left\{Z_{t+1} \le \frac{x}{\alpha_0}\right\} \mathbb{I}\left(x \ge \frac{\alpha_1}{\alpha_0}x_t\right) \mathbb{I}\left(x \ge \frac{\alpha_2}{\alpha_0}x_{t-1}\right) \\ &= \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_0}x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_2}{\alpha_0}x_{t-1}. \end{cases} \end{aligned} \tag{4.99}$$

On the other hand, if $Z_{t-2} = \frac{x_t}{\alpha_2}$ then

$$\begin{aligned} F_{t+1}^{(2,0)}(x) &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0},\ Z_t \le \frac{x}{\alpha_1},\ Z_{t-1} \le \frac{x}{\alpha_2} \,\middle|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_{t-1}}{\alpha_0},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0} \right\} P\left\{ Z_t \le \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} \mathbb{I}\left(x \ge \frac{\alpha_2}{\alpha_0} x_{t-1} \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_2}{\alpha_0} x_{t-1}, \end{cases} \end{aligned} \tag{4.100}$$

which follows because $P\left\{ Z_t \le \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} = 1$ for $x \ge \frac{\alpha_2}{\alpha_0} x_{t-1}$. Therefore

$$F_{t+1}(x) = \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \tag{4.101}$$

Case 3.2: If $\alpha_1/\alpha_0 < x_t/x_{t-1} < \alpha_2/\alpha_1$ and $x_{t-1}/x_{t-2} \le \alpha_2/\alpha_1$ then $Z_{t-1} < \frac{x_t}{\alpha_1}$, and $Z_{t-2} < \frac{x_{t-1}}{\alpha_1}$. However, X_t can be either $\alpha_0 Z_t$ or $\alpha_2 Z_{t-2}$, and X_{t-1} can be either $\alpha_0 Z_{t-1}$ or $\alpha_2 Z_{t-3}$.

First consider the probability of each of the two possible cases for X_{t-1}. For infinitesimally small δ_{t-1}, we have

$$\begin{aligned} &P\left\{ Z_{t-1} \in \frac{(x_{t-1}, x_{t-1} + \delta_{t-1})}{\alpha_0},\ Z_{t-3} < \frac{x_{t-1}}{\alpha_2} \,\middle|\, \mathcal{F}_t \right\} \\ &\quad = P\left\{ Z_{t-1} \in \frac{(x_{t-1}, x_{t-1} + \delta_{t-1})}{\alpha_0} \,\middle|\, Z_{t-1} < \frac{x_t}{\alpha_1} \right\} P\left\{ Z_{t-3} < \frac{x_{t-1}}{\alpha_2} \,\middle|\, Z_{t-3} \le \frac{x_{t-2}}{\alpha_1} \right\} \\ &\quad = \frac{\alpha_0 \delta_{t-1}}{x_{t-1}^2} \exp\left\{ -\frac{\alpha_0 + \alpha_2}{x_{t-1}} + \frac{\alpha_1}{x_t} + \frac{\alpha_1}{x_{t-2}} \right\}, \end{aligned} \tag{4.102}$$

and

$$\begin{aligned} &P\left\{ Z_{t-1} < \frac{x_{t-1}}{\alpha_0},\ Z_{t-3} \in \frac{(x_{t-1}, x_{t-1} + \delta_{t-1})}{\alpha_2} \,\middle|\, \mathcal{F}_t \right\} \\ &\qquad = \frac{\alpha_2 \delta_{t-1}}{x_{t-1}^2} \exp\left\{ -\frac{\alpha_0 + \alpha_2}{x_{t-1}} + \frac{\alpha_1}{x_t} + \frac{\alpha_1}{x_{t-2}} \right\}. \end{aligned} \tag{4.103}$$

Then we have

$$P\{\alpha_0 Z_{t-1} > \alpha_2 Z_{t-3} \mid \mathcal{F}_t\} = \frac{\alpha_0}{\alpha_0 + \alpha_2}. \tag{4.104}$$

Similar calculations give the relative frequencies of each of the two possible cases for X_t. In other words, we have

$$P\{\alpha_0 Z_t > \alpha_2 Z_{t-2} \mid \mathcal{F}_t\} = \frac{\alpha_0}{\alpha_0 + \alpha_2}. \tag{4.105}$$

On the other hand, if $Z_{t-1} = \frac{x_{t-1}}{\alpha_0}$ then we obtain equation (4.101). Thus we can write

$$\begin{aligned} F_{t+1}(x) = &\frac{\alpha_0}{\alpha_0 + \alpha_2} \exp\left\{-\frac{\alpha_0}{x}\right\} \mathbb{I}\left(x \geq \frac{\alpha_2}{\alpha_0} x_{t-1}\right) \\ &+ \frac{\alpha_2}{\alpha_0 + \alpha_2}\left[\frac{\alpha_0}{\alpha_0 + \alpha_2} F_{t+1}^{(0,2)}(x) + \frac{\alpha_2}{\alpha_0 + \alpha_2} F_{t+1}^{(2,2)}(x)\right], \end{aligned} \tag{4.106}$$

where $F_{t+1}^{(i,j)}(x)$ is defined in (4.98).

Next, we obtain explicit expressions for $F_{t+1}^{(0,2)}(x)$ and $F_{t+1}^{(2,2)}(x)$. Note that $\frac{x_t}{x_{t-1}} < \frac{\alpha_2}{\alpha_1}$ gives $\frac{\alpha_1}{\alpha_0} x_t < \frac{\alpha_2}{\alpha_0} x_{t-1}$, then

$$\begin{aligned} F_{t+1}^{(0,2)}(x) &= P\left\{ X_{t+1} \leq x \mid Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_{t-1}}{\alpha_0},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0} x_t\right) P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0} \right\} \\ &= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_0}{x_{t-1}}\right\}, & \frac{\alpha_1}{\alpha_0} x_t \leq x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \end{aligned} \tag{4.107}$$

Similarly,

$$\begin{aligned} F_{t+1}^{(2,2)}(x) &= P\left\{ X_{t+1} \leq x \mid Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_{t-1}}{\alpha_0},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} P\left\{ Z_t \leq \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0} \right\} \end{aligned}$$

$$= \begin{cases} \exp\left\{-\frac{1}{x} + \frac{\alpha_0}{x_t} + \frac{\alpha_0}{x_{t-1}}\right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_0}{x_{t-1}}\right\}, & \frac{\alpha_1}{\alpha_0} x_t \le x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \tag{4.108}$$

Substituting (4.107) and (4.108) into (4.106) gives

$$F_{t+1}(x) = \begin{cases} \left(\frac{\alpha_2}{\alpha_0+\alpha_2}\right)^2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0}{x_t} + \frac{\alpha_0}{x_{t-1}}\right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \frac{\alpha_2}{\alpha_0+\alpha_2} \exp\left\{-\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_0}{x_{t-1}}\right\}, & \frac{\alpha_1}{\alpha_0} x_t \le x < \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \tag{4.109}$$

Case 4: If $x_t/x_{t-1} = \alpha_1/\alpha_0$ then $Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$, therefore

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0},\ Z_t \le \frac{x}{\alpha_1},\ Z_{t-1} \le \frac{x}{\alpha_2} \,\middle|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_t}{\alpha_1} \right\} \\ &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0} \right\} P\left\{ Z_t \le \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} \mathbb{I}\left(x \ge \frac{\alpha_2}{\alpha_1} x_t \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t = \frac{\alpha_2}{\alpha_0} x_{t-1}, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \ge \frac{\alpha_2}{\alpha_1} x_t = \frac{\alpha_2}{\alpha_0} x_{t-1}. \end{cases} \end{aligned} \tag{4.110}$$

Case 5: If $x_t/x_{t-1} < \alpha_1/\alpha_0$ then $Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$ and $Z_{t-2} \le \min\left(\frac{x_{t-1}}{\alpha_1}, \frac{x_{t-2}}{\alpha_0}\right)$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$.

First note that $x_t/x_{t-1} < \alpha_1/\alpha_0$ implies $x_{t-1}/x_{t-2} \le \alpha_2/\alpha_1$. Second, we claim that $\frac{\alpha_1}{\alpha_0} x_{t-1} < \min\left(\frac{\alpha_2}{\alpha_1} x_{t-1}, \frac{\alpha_2}{\alpha_0} x_{t-2}\right)$. This follows because if $\alpha_1/\alpha_0 < x_{t-1}/x_{t-2} \le \alpha_2/\alpha_1$ then $\frac{\alpha_1}{\alpha_0} x_{t-1} < \frac{\alpha_2}{\alpha_0} x_{t-2}$. On the other hand, if $x_{t-1}/x_{t-2} \le \alpha_1/\alpha_0$ then $\frac{x_{t-1}}{\alpha_1} < \frac{x_{t-2}}{\alpha_0}$, so $\frac{\alpha_1}{\alpha_0} x_{t-1} < \frac{\alpha_2}{\alpha_1} x_{t-1}$. Denote $x^* = \min\left(\frac{x_{t-1}}{\alpha_1}, \frac{x_{t-2}}{\alpha_0}\right)$. Then for infinitesimally small δ_t,

$$\begin{aligned} P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, \mathcal{F}_t \right\} \\ = P\left\{ Z_t \in \frac{(x_t, x_t+\delta_t)}{\alpha_0} \right\} P\left\{ Z_{t-1} < \frac{x_t}{\alpha_1} \,\middle|\, Z_{t-1} < \frac{x_{t-1}}{\alpha_0} \right\} \\ \times P\left\{ Z_{t-2} < \frac{x_t}{\alpha_2} \,\middle|\, Z_{t-2} \le x^* \right\} \end{aligned}$$

$$= \frac{\alpha_0 \delta_t}{x_t^2} \exp\left\{ -\frac{1}{x_t} + \frac{\alpha_0}{x_{t-1}} + \frac{1}{x^*} \right\}, \tag{4.111}$$

By the same argument used in previous cases it follows that

$$P\left\{ \alpha_i Z_{t-i} > \alpha_k Z_{t-k},\ k \neq i \;\middle|\; Z_{t-3} = \frac{x_{t-1}}{\alpha_2},\ \mathcal{F}_t \right\} = \alpha_i, \qquad i = 0, 1, 2. \tag{4.112}$$

Thus we can write

$$F_{t+1}(x) = \alpha_0 F_{t+1}^{(0,2)}(x) + \alpha_1 F_{t+1}^{(1,2)}(x) + \alpha_2 F_{t+1}^{(2,2)}(x), \tag{4.113}$$

where $F_{t+1}^{(i,j)}(x)$ is defined in (4.98).

We now consider $F_{t+1}^{(0,2)}(x)$, $F_{t+1}^{(1,2)}(x)$, and $F_{t+1}^{(2,2)}(x)$ in detail:

- If $Z_t = \frac{x_t}{\alpha_0}$ then $Z_{t-1} < \frac{x_t}{\alpha_1} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0} \right)$ and

$$\begin{aligned} F_{t+1}^{(0,2)}(x) &= P\left\{ X_{t+1} \leq x \,|\, Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0} x_t \right) P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \;\middle|\; Z_{t-1} < \frac{x_t}{\alpha_1} \right\} \\ &= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0 + \alpha_2}{x} + \frac{\alpha_1}{x_t} \right\}, & \frac{\alpha_1}{\alpha_0} x_t \leq x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_2}{\alpha_1} x_t. \end{cases} \end{aligned} \tag{4.114}$$

- If $Z_{t-1} = \frac{x_t}{\alpha_1}$ then

$$\begin{aligned} F_{t+1}^{(1,2)}(x) &= P\left\{ X_{t+1} \leq x \,|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_t}{\alpha_1},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} P\left\{ Z_t \leq \frac{x}{\alpha_1} \;\middle|\; Z_t < \frac{x_t}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_2}{\alpha_1} x_t \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \end{aligned} \tag{4.115}$$

- If $Z_{t-2} = \frac{x_t}{\alpha_2}$ then $Z_t < \frac{x_t}{\alpha_0}$, $Z_{t-1} < \frac{x_t}{\alpha_1}$, and

$$\begin{aligned} F_{t+1}^{(2,2)}(x) &= P\left\{ X_{t+1} \le x \,|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1},\ \mathcal{F}_t \right\} \\ &= P\left\{ Z_{t+1} \le \frac{x}{\alpha_0} \right\} P\left\{ Z_t \le \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} P\left\{ Z_{t-1} \le \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_t}{\alpha_1} \right\} \\ &= \begin{cases} \exp\left\{ -\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t} \right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_1}{x_t} \right\}, & \frac{\alpha_1}{\alpha_0} x_t \le x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_2}{\alpha_1} x_t. \end{cases} \end{aligned} \tag{4.116}$$

Substituting (4.114), (4.115), and (4.116) into (4.113) gives

$$F_{t+1}(x) = \begin{cases} \alpha_2 \exp\left\{ -\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t} \right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ (\alpha_0+\alpha_2) \exp\left\{ -\frac{\alpha_0+\alpha_2}{x} + \frac{\alpha_1}{x_t} \right\}, & \frac{\alpha_1}{\alpha_0} x_t \le x < \frac{\alpha_2}{\alpha_1} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \ge \frac{\alpha_2}{\alpha_1} x_t. \end{cases} \tag{4.117}$$

□

Proof of Proposition 4.2.3.

Assume $\alpha_2/\alpha_1 = \alpha_1/\alpha_0$. Consider all possible scenarios for the ratios X_t/X_{t-1} and X_{t-1}/X_{t-2} as shown in Table 4.3. We deduce five cases as summarized in Table 4.5, and consider each of these cases in detail.

Case	x_t/x_{t-1}	x_{t-1}/x_{t-2}	Deduction
1	$> \alpha_1/\alpha_0$		$Z_t = \frac{x_t}{\alpha_0}$ and $Z_{t-1} \le \frac{x_{t-1}}{\alpha_0}$.
2.1	$= \alpha_1/\alpha_0$	$> \alpha_1/\alpha_0$	$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$.
2.2	$= \alpha_1/\alpha_0$	$= \alpha_1/\alpha_0$	$Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} < \frac{x_t}{\alpha_1}$.
2.3	$= \alpha_1/\alpha_0$	$< \alpha_1/\alpha_0$	$Z_t < \frac{x_t}{\alpha_0}$. However, we can have either $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$ or $Z_{t-2} = \frac{x_t}{\alpha_2} = \frac{x_{t-1}}{\alpha_1}$.
3	$< \alpha_1/\alpha_0$		$Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$.

Table 4.5: *Deduced Cases for the Proof of Proposition 4.2.3.*

Case 1: If $x_t/x_{t-1} > \alpha_1/\alpha_0$, then $Z_t = \frac{x_t}{\alpha_0}$ and $Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$. Note that $\frac{x_t}{x_{t-1}} > \frac{\alpha_1}{\alpha_0} = \frac{\alpha_2}{\alpha_1} \iff \frac{\alpha_1}{\alpha_0}x_t > \frac{\alpha_2}{\alpha_0}x_{t-1}$, and $P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} = 1$ for $x \geq \frac{\alpha_1}{\alpha_0}x_t$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0},\ Z_t \leq \frac{x}{\alpha_1},\ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_t = \frac{x_t}{\alpha_0},\ Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0}x_t \right) P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} \leq \frac{x_{t-1}}{\alpha_0} \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0}x_t \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \end{aligned} \tag{4.118}$$

Case 2.1: If $x_t/x_{t-1} = \alpha_1/\alpha_0$ and $x_{t-1}/x_{t-2} > \alpha_1/\alpha_0$ then $Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0},\ Z_t \leq \frac{x}{\alpha_1},\ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} = \frac{x_t}{\alpha_1} \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} P\left\{ Z_t \leq \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} \mathbb{I}\left(x \geq \frac{\alpha_1}{\alpha_0}x_t \right) \\ &= \begin{cases} 0, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \end{aligned} \tag{4.119}$$

Case 2.2: If $x_t/x_{t-1} = \alpha_1/\alpha_0$ and $x_{t-1}/x_{t-2} = \alpha_1/\alpha_0$ then $Z_t < \frac{x_t}{\alpha_0}$ and $Z_{t-1} < \frac{x_t}{\alpha_1} = \min\left(\frac{x_t}{\alpha_1}, \frac{x_{t-1}}{\alpha_0}\right)$. Then

$$\begin{aligned} F_{t+1}(x) &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0},\ Z_t \leq \frac{x}{\alpha_1},\ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_t < \frac{x_t}{\alpha_0},\ Z_{t-1} < \frac{x_t}{\alpha_1} \right\} \\ &= P\left\{ Z_{t+1} \leq \frac{x}{\alpha_0} \right\} P\left\{ Z_t \leq \frac{x}{\alpha_1} \,\middle|\, Z_t < \frac{x_t}{\alpha_0} \right\} P\left\{ Z_{t-1} \leq \frac{x}{\alpha_2} \,\middle|\, Z_{t-1} < \frac{x_t}{\alpha_1} \right\} \\ &= \begin{cases} \exp\left\{ -\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t} \right\}, & x < \frac{\alpha_1}{\alpha_0}x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0}x_t. \end{cases} \end{aligned} \tag{4.120}$$

Case 2.3: If $x_t/x_{t-1} = \alpha_1/\alpha_0$ and $x_{t-1}/x_{t-2} < \alpha_1/\alpha_0$ then $Z_t < \frac{x_t}{\alpha_0}$. However, we can

have either $Z_{t-1} = \frac{x_t}{\alpha_1} = \frac{x_{t-1}}{\alpha_0}$ (Case 2.1) or $Z_{t-2} = \frac{x_t}{\alpha_2} = \frac{x_{t-1}}{\alpha_1}$ (Case 2.2).

Since $Z_{t-3} < \frac{x_{t-1}}{\alpha_2}$ we only need to consider, for infinitesimally small δ_{t-1},

$$\begin{aligned} &P\left\{ Z_{t-1} \in \frac{(x_{t-1}, x_{t-1}+\delta_{t-1})}{\alpha_0},\ Z_{t-2} < \frac{x_{t-1}}{\alpha_1} \,\middle|\, \mathcal{F}_t \right\} \\ &\quad = P\left\{ Z_{t-1} \in \frac{(x_{t-1}, x_{t-1}+\delta_{t-1})}{\alpha_0},\ Z_{t-2} < \frac{x_{t-1}}{\alpha_1} \,\middle|\, Z_{t-2} < \frac{x_{t-2}}{\alpha_0} \right\} \\ &\quad = \frac{\alpha_0 \delta_{t-1}}{x_{t-1}^2} \exp\left\{ -\frac{\alpha_0+\alpha_1}{x_{t-1}} + \frac{\alpha_0}{x_{t-2}} \right\}, \end{aligned} \tag{4.121}$$

and

$$\begin{aligned} &P\left\{ Z_{t-1} < \frac{x_{t-1}}{\alpha_1},\ Z_{t-2} \in \frac{(x_{t-1}, x_{t-1}+\delta_{t-1})}{\alpha_0} \,\middle|\, \mathcal{F}_t \right\} \\ &\qquad = \frac{\alpha_1 \delta_{t-1}}{x_{t-1}^2} \exp\left\{ -\frac{\alpha_0+\alpha_1}{x_{t-1}} + \frac{\alpha_0}{x_{t-2}} \right\}. \end{aligned} \tag{4.122}$$

Then it follows that

$$P\left\{ \alpha_0 Z_{t-1} > \alpha_1 Z_{t-2} \,|\, \mathcal{F}_t \right\} = \frac{\alpha_0}{\alpha_0+\alpha_1}. \tag{4.123}$$

By combining (4.119), (4.120) and (4.123) we obtain

$$\begin{aligned} F_{t+1}(x) &= \frac{\alpha_0}{\alpha_0+\alpha_1} F_{t+1}^{(1,0)}(x) + \frac{\alpha_1}{\alpha_0+\alpha_1} F_{t+1}^{(2,1)}(x) \\ &= \begin{cases} \frac{\alpha_1}{\alpha_0+\alpha_1} \exp\left\{ -\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t} \right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{ -\frac{\alpha_0}{x} \right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \end{aligned} \tag{4.124}$$

Case 3: If $x_t/x_{t-1} < \alpha_1/\alpha_0$ then $Z_{t-3} = \frac{x_{t-1}}{\alpha_2}$, $Z_{t-2} < \frac{x_{t-1}}{\alpha_1} = \min\left(\frac{x_{t-1}}{\alpha_1}, \frac{x_{t-2}}{\alpha_0} \right)$, and $Z_{t-1} < \frac{x_{t-1}}{\alpha_0}$. However, X_t can be either $\alpha_0 Z_t$, $\alpha_1 Z_{t-1}$, or $\alpha_2 Z_{t-2}$. Then as shown in Case 5 of the proof of Proposition 4.2.1 we have

$$P\left\{ \alpha_i Z_{t-i} > \alpha_k Z_{t-k},\ k \neq i \,\middle|\, Z_{t-3} = \frac{x_{t-1}}{\alpha_2},\ \mathcal{F}_t \right\} = \alpha_i, \qquad i = 0, 1, 2. \tag{4.125}$$

Thus we can write

$$F_{t+1}(x) = \alpha_0 F_{t+1}^{(0)}(x) + \alpha_1 F_{t+1}^{(1)}(x) + \alpha_2 F_{t+1}^{(2)}(x), \tag{4.126}$$

where $F_{t+1}^{(i)}(x)$ is defined in (4.85).

On the other hand, $F_{t+1}^{(0)}(x)$ is given by (4.118), $F_{t+1}^{(1)}(x)$ is given by (4.119), and $F_{t+1}^{(2)}(x)$ is given by (4.120). Substituting all of these into (4.126), gives

$$F_{t+1}(x) = \begin{cases} \alpha_2 \exp\left\{-\frac{1}{x} + \frac{\alpha_0+\alpha_1}{x_t}\right\}, & x < \frac{\alpha_1}{\alpha_0} x_t, \\ \exp\left\{-\frac{\alpha_0}{x}\right\}, & x \geq \frac{\alpha_1}{\alpha_0} x_t. \end{cases} \tag{4.127}$$

□

CHAPTER 5

State-Space Models and Particle Filters

Many time series models can be represented as a *state-space* model, where there is an underlying unobserved *state process* evolving according to a known stochastic law, and an *observation process* that is dependent on the state process. The main goal is to estimate the unobserved state process in terms of the observed process. There are three estimation problems associated with the state-space model: prediction, filtering, and smoothing. In this Chapter we restrict our exposition to the filtering problem.

It is well known that when the state and the observation equations are both linear and Gaussian then the Kalman recursions (Kalman, 1960) can be used to obtain the optimal filter. The literature on linear state-space models is extensive, for instance Harvey (1989) and West and Harrison (1997). A good introduction is Chapter 12 of Brockwell and Davis (1993).

However, in the presence of nonlinearities or when the process is non-Gaussian there is no closed-form solution for the optimal filter and approximations have to be made. Particle filters are a class of simulation filters that approximate the optimal filter by a discrete set of particles. Finding the probabilistic and statistical properties of particle filters is a topic of current research. In this Chapter we describe filtering and particle methods for filtering. Our exposition is mainly based on Doucet, De Freitas and Gordon (2001).

5.1 Filtering

Let $\{x_t\}$ be the unobserved state process, and let $\{y_t\}$ be the observed process (data). The standard formulation of the state-space model assumes that the state is first-order Markov, i.e. x_t is conditionally independent of past states and observations given x_{t-1}.

However, some of the models considered in this work are not Markov, hence we need to keep both cases in mind. We introduce the notation $x_{0:t} = (x_0, x_1, \ldots, x_t)$ and $y_{1:t} = (y_1, \ldots, y_t)$. In general, the state-space model we consider is

$$f(x_0), \tag{5.1}$$

$$f(x_t \mid x_{0:t-1}) \quad \text{for } t \geq 1 \quad \text{(state transition equation)}, \tag{5.2}$$

$$f(y_t \mid x_t) \quad \text{for } t \geq 1 \quad \text{(observation equation)}. \tag{5.3}$$

In this model each y_t is conditionally independent of past states and observations given the current state x_t, and the transition equation of the states depends only on past values of the state. If we assume the state is first-order Markov then we substitute (5.2) with $f(x_t \mid x_{t-1})$.

We are interested in the joint distribution of the states given the observations

$$f(x_{0:t} \mid y_{1:t}) = \frac{f(y_{1:t} \mid x_{0:t}) f(x_{0:t})}{f(y_{1:t})}, \tag{5.4}$$

the marginal $f(x_t \mid y_{1:t})$, and expectations of the form

$$\mathbb{E}(h(x_{0:t}) \mid y_{1:t}) = \int h(x_{0:t}) f(x_{0:t} \mid y_{1:t}) \, \mathrm{d}x_{0:t}, \tag{5.5}$$

for an integrable function h.

Recursive Formula for $f(x_{0:t+1} \mid y_{1:t+1})$

It is possible to write a recursive formula for the posterior distribution $f(x_{0:t+1} \mid y_{1:t+1})$. Assuming the state is Markov we have

$$\begin{aligned}
f(x_{0:t+1} \mid y_{1:t+1}) &= \frac{f(y_{1:t+1} \mid x_{0:t+1}) f(x_{0:t+1})}{f(y_{1:t+1})} \\
&= \frac{f(y_{t+1} \mid x_{0:t+1}, y_{1:t}) f(y_{1:t} \mid x_{0:t+1}) f(x_{t+1} \mid x_{0:t}) f(x_{0:t})}{f(y_{t+1} \mid y_{1:t}) f(y_{1:t})} \\
&= \frac{f(y_{t+1} \mid x_{t+1}) f(y_{1:t} \mid x_{0:t}) f(x_{t+1} \mid x_t) f(x_{0:t})}{f(y_{t+1} \mid y_{1:t}) f(y_{1:t})} \\
&= f(x_{0:t} \mid y_{1:t}) \frac{f(y_{t+1} \mid x_{t+1}) f(x_{t+1} \mid x_t)}{f(y_{t+1} \mid y_{1:t})},
\end{aligned} \tag{5.6}$$

which follows by independence of the y_t's and the fact that y_{t+1} only depends on x_{t+1}. On the other hand, if the process is not Markov, but the observations y_t are still conditionally independent given x_t, i.e.

$$f\left(y_{1:t} \mid x_{0:t}\right) = \prod_{s=1}^{t} f\left(y_s \mid x_s\right), \tag{5.7}$$

then we have

$$f\left(x_{0:t+1} \mid y_{1:t+1}\right) = f\left(x_{0:t} \mid y_{1:t}\right) \frac{f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid x_{0:t}\right)}{f\left(y_{t+1} \mid y_{1:t}\right)}. \tag{5.8}$$

Recursive Formula for $f\left(x_{t+1} \mid y_{1:t+1}\right)$

Usually the interest is on the marginal $f\left(x_{t+1} \mid y_{1:t+1}\right)$, which is known as the *filtering distribution*. Recalling the identity

$$f\left(x \mid y\right) = \int \frac{f\left(x, y, z\right)}{f\left(y, z\right)} \frac{f\left(y, z\right)}{f\left(y\right)} \mathrm{d}z = \int f\left(x \mid y, z\right) f\left(z \mid y\right) \mathrm{d}z, \tag{5.9}$$

the *prediction equation* when the state is Markov is

$$\begin{aligned} f\left(x_{t+1} \mid y_{1:t}\right) &= \int f\left(x_{t+1} \mid x_t, y_{1:t}\right) f\left(x_t \mid y_{1:t}\right) \mathrm{d}x_t \\ &= \int f\left(x_{t+1} \mid x_t\right) f\left(x_t \mid y_{1:t}\right) \mathrm{d}x_t. \end{aligned} \tag{5.10}$$

If the state is not Markov the prediction equation is

$$\begin{aligned} f\left(x_{t+1} \mid y_{1:t}\right) &= \int f\left(x_{t+1} \mid x_{0:t}, y_{1:t}\right) f\left(x_{0:t} \mid y_{1:t}\right) \mathrm{d}x_{0:t} \\ &= \int f\left(x_{t+1} \mid x_{0:t}\right) f\left(x_{0:t} \mid y_{1:t}\right) \mathrm{d}x_{0:t}. \end{aligned} \tag{5.11}$$

Once the new observation y_{t+1} has arrived, the prediction equation is updated via

$$\begin{aligned} f\left(x_{t+1} \mid y_{1:t+1}\right) &= \frac{f\left(x_{t+1}, y_{1:t+1}\right)}{f\left(y_{1:t+1}\right)} \\ &= \frac{f\left(y_{t+1} \mid x_{t+1}, y_{1:t}\right) f\left(x_{t+1} \mid y_{1:t}\right) f\left(y_{1:t}\right)}{f\left(y_{1:t+1}\right)} \end{aligned}$$

$$= \frac{f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right)}{f\left(y_{t+1} \mid y_{1:t}\right)}, \tag{5.12}$$

and since

$$\begin{aligned} f\left(y_{t+1} \mid y_{1:t}\right) &= \int f\left(y_{t+1} \mid x_{t+1}, y_{1:t}\right) f\left(x_{t+1} \mid y_{1:t}\right) \mathrm{d}x_{t+1} \\ &= \int f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right) \mathrm{d}x_{t+1}, \end{aligned} \tag{5.13}$$

the *updating equation* is

$$f\left(x_{t+1} \mid y_{1:t+1}\right) = \frac{f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right)}{\int f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right) \mathrm{d}x_{t+1}}. \tag{5.14}$$

Therefore the marginal $f\left(x_{t+1} \mid y_{1:t+1}\right)$ satisfies the following two-step recursion:

Prediction:
$$f\left(x_{t+1} \mid y_{1:t}\right) = \int f\left(x_{t+1} \mid x_{0:t}\right) f\left(x_{0:t} \mid y_{1:t}\right) \mathrm{d}x_{0:t} \tag{5.15}$$

Updating:
$$f\left(x_{t+1} \mid y_{1:t+1}\right) = \frac{f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right)}{\int f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right) \mathrm{d}x_{t+1}} \tag{5.16}$$

5.2 Particle Filters

When the state-space is nonlinear and non-Gaussian, there is no exact solution for the filtering distribution $f\left(x_{t+1} \mid y_{1:t+1}\right)$. *Particle filters* (also known as *sequential Monte Carlo methods*) are simulation based filters that can be used to solve this problem. There is now a large literature on particle filter methods. A comprehensive overview of particle filters can be found in Doucet *et al.* (2001). Good introductory references are Gordon, Salmond and Smith (1993), Kitagawa (1996), and Arulampalam *et al.* (2002).

To simplify the exposition, in what follows we restrict ourselves to the Markov case, though the methods can still be applied in the more general setting. Following Pitt and Shephard (1999a, 2001), filtering can be thought of as the repeated application of the iteration

$$f\left(x_{t+1} \mid y_{1:t+1}\right) \propto f\left(y_{t+1} \mid x_{t+1}\right) \int f\left(x_{t+1} \mid x_t\right) f\left(x_t \mid y_{1:t}\right) \mathrm{d}x_t. \tag{5.17}$$

Suppose that at time t, the filtering distribution $f(x_t \mid y_{1:t})$ is approximated by a discrete sample $\left\{x_t^{(i)},\ i = 1, \ldots, N\right\}$ (called particles) with probability masses $\left\{w_t^{(i)},\ i = 1, \ldots, N\right\}$. Particle filters are the class of simulation filters that recursively approximate the filtering distribution (5.17) by

$$\hat{f}(x_{t+1} \mid y_{1:t+1}) \propto f(y_{t+1} \mid x_{t+1}) \sum_{i=1}^{N} f\left(x_{t+1} \,\middle|\, x_t^{(i)}\right) w_t^{(i)}. \tag{5.18}$$

The basic particle filter algorithm, known as the bootstrap filter, as described in Doucet *et al.* (2001) is presented in Algorithm 5.2.1.

Algorithm 5.2.1 (Bootstrap Filter).

1. *Initialization*, $t = 0$
   - For $i = 1, \ldots, N$, sample $x_0^{(i)} \sim f(x_0)$.
2. *Importance Sampling Step*
   - For $i = 1, \ldots, N$, sample $\tilde{x}_{t+1}^{(i)} \sim f\left(x_{t+1} \,\middle|\, x_t^{(i)}\right)$.
   - For $i = 1, \ldots, N$, evaluate the importance weights

$$\widetilde{w}_{t+1}^{(i)} = f\left(y_{t+1} \,\middle|\, \tilde{x}_{t+1}^{(i)}\right), \qquad w_{t+1}^{(i)} = \frac{\widetilde{w}_{t+1}^{(i)}}{\sum_{i=1}^{N} \widetilde{w}_{t+1}^{(i)}}.$$

3. *Selection Step*
   - Resample with replacement N particles $\left\{x_{t+1}^{(i)},\ i = 1, \ldots, N\right\}$ from the set $\left\{\tilde{x}_{t+1}^{(i)},\ i = 1, \ldots, N\right\}$ according to the importance weights.
   - Set $t \leftarrow t + 1$ and go to step 2.

The algorithm described above is an application of the weighted bootstrap algorithm (Smith and Gelfand, 1992) to obtain a sample that is approximately distributed according to $f(x_{t+1} \mid y_{1:t+1})$. This is a variant of the well known bootstrap (Efron, 1982). Rubin (1988) refers to this procedure as sampling/importance resampling (SIR). The weighted bootstrap algorithm is described in §B.2 for reference.

The weighted bootstrap is not an essential assumption for particle filters to work. We can also use rejection sampling or Markov chain Monte Carlo (MCMC), see Pitt and Shep-

hard (1999a, p. 591).

5.3 Auxiliary Variable Based Particle Filters

The basic particle algorithm often fails in practice because the presence of an outlier makes the importance weights very unevenly distributed, so after a few time steps there are only few different particles, and eventually the particles degenerate to a single value. *Auxiliary particle filters* were introduced by Pitt and Shephard (1999a) to improve the sampling from the target distribution. Pitt and Shephard (1999a), define

$$f\left(x_{t+1}, J \mid y_{1:t+1}\right) \propto f\left(y_{t+1} \mid x_{t+1}\right) f\left(x_{t+1} \left| x_t^{(J)}\right.\right) w_t^{(J)}, \qquad J = 1, \ldots, N. \tag{5.19}$$

The idea is to sample from the joint density (5.19) where $J \in \{1, \ldots, N\}$ is an index from the mixture (5.18), and then discard the index J to produce a sample from the empirical filtering density (5.18) as required. The justification of this approach is that by integrating out J in (5.19) we have

$$f\left(x_{t+1} \mid y_{1:t+1}\right) \propto f\left(y_{t+1} \mid x_{t+1}\right) \sum_{J=1}^{N} f\left(x_{t+1} \left| x_t^{(J)}\right.\right) w_t^{(J)}. \tag{5.20}$$

The index J is called an auxiliary variable because it is present simply to aid the task of the simulation. Generic particle filters of this type are called *auxiliary particle filters*.

It is possible to sample from $f\left(x_{t+1}, J \mid y_{1:t+1}\right)$ using weighted bootstrap, rejection sampling or MCMC. Using the weighted bootstrap we need to make N proposals

$$\left(x_{t+1}^{(i)}, J^{(i)}\right) \sim g\left(x_{t+1}, J \mid y_{1:t+1}\right), \qquad i = 1, \ldots, N$$

and then construct resampling weights

$$w_{t+1}^{(i)} \propto \frac{f\left(y_{t+1} \left| x_{t+1}^{(i)}\right.\right) f\left(x_{t+1}^{(i)} \left| x_t^{\left(J^{(i)}\right)}\right.\right)}{g\left(x_{t+1}^{(i)}, J^{(i)} \left| y_{1:t+1}\right.\right)}, \qquad i = 1, \ldots, N.$$

A convenient generic suggestion for the choice of $g(\cdot)$ given by Pitt and Shephard

(1999a) is to approximate (5.19) by

$$g\left(x_{t+1}, J \mid y_{1:t+1}\right) \propto f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}\right) f\left(x_{t+1} \,\middle|\, x_t^{(J)}\right) w_t^{(J)}, \qquad J = 1, \ldots, N,$$

where $\nu_{t+1}^{(J)}$ is the mean, the mode, a draw, or some other likely value associated with the density of $\left(x_{t+1} \,\middle|\, x_t^{(J)}\right)$. Note that

$$\begin{aligned} g\left(J \mid y_{1:t+1}\right) &\propto w_t^{(J)} \int f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}\right) f\left(x_{t+1} \,\middle|\, x_t^{(J)}\right) \mathrm{d}x_{t+1} \\ &= w_t^{(J)} f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}\right). \end{aligned}$$

So we can sample first from $g\left(J \mid y_{1:t+1}\right)$ and then given J, sample from $f\left(x_{t+1} \,\middle|\, x_t^{(J)}\right)$. In this case

$$g\left(x_{t+1}, J \mid y_{1:t+1}\right) \propto \underbrace{g\left(x_{t+1} \mid J, y_{1:t+1}\right)}_{f\left(x_{t+1} \mid x_t^{(J)}\right)} \underbrace{g\left(J \mid y_{1:t+1}\right)}_{w_t^{(J)} f\left(y_{t+1} \mid \nu_{t+1}^{(J)}\right)}.$$

The corresponding weights become

$$\begin{aligned} w_t^{(i)} &\propto \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}\right) f\left(x_{t+1}^{(i)} \,\middle|\, x_t^{\left(J^{(i)}\right)}\right)}{g\left(x_{t+1}^{(i)}, J^{(i)} \,\middle|\, y_{1:t+1}\right)} \\ &= \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}\right)}{f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{\left(J^{(i)}\right)}\right)}, \qquad i = 1, \ldots, N. \end{aligned}$$

5.4 Particle Filters and Unknown Parameters

Usually there are unknown parameters in the state-space formulation. Particle filtering algorithms dealing simultaneously with both fixed model parameters and state variables have been proposed in the literature. However, estimation of static parameters with particle filters is not recommended, as pointed out in Crisan and Doucet (2002, Remark 2, p. 745).

Doucet and Tadić (2003) propose particle methods for maximum likelihood estimation of fixed parameters using gradient-type algorithms. This approach requires calculation of derivatives of the state transition distribution.

Liu and West (2001) consider the following state-space model:

$$f(x_0, \theta), \tag{5.21}$$

$$f(x_t \mid x_{t-1}, \theta) \quad \text{for } t \geq 1 \quad \text{(state transition equation)}, \tag{5.22}$$

$$f(y_t \mid x_t, \theta) \quad \text{for } t \geq 1 \quad \text{(observation equation)}. \tag{5.23}$$

Each y_t is conditionally independent of past states and observations given the current state x_t, and the parameter θ; and x_t is conditionally independent of past states and observations given x_{t-1} and θ. In this model the state vector at time t is a combined (augmented) space (x_t, θ). By Bayes' theorem we have

$$\begin{aligned} f(x_{t+1}, \theta \mid y_{1:t+1}) &= \frac{f(y_{1:t+1}, x_{t+1}, \theta)}{f(y_{1:t+1})} \\ &= \frac{f(y_{t+1} \mid x_{t+1}, \theta, y_{1:t})\, f(x_{t+1} \mid \theta, y_{1:t})\, f(\theta \mid y_{1:t})\, f(y_{1:t})}{f(y_{1:t+1})} \\ &\propto f(y_{t+1} \mid x_{t+1}, \theta)\, f(x_{t+1} \mid \theta, y_{1:t})\, f(\theta \mid y_{1:t}). \end{aligned}$$

Liu and West (2001) proposed an auxiliary particle filter on the augmented space. The distribution $f(x_{t+1}, \theta \mid y_{1:t+1})$ is approximated by particles

$$\left(x_{t+1}^{(i)}, \theta_{t+1}^{(i)}\right), \qquad i = 1, \ldots, N,$$

with corresponding weights $w_{t+1}^{(i)}$, $i = 1, \ldots, N$. The parameter θ is static, so the $t+1$ suffix on the θ samples only indicate that they are from the time $t+1$ posterior, not that θ is time-varying.

5.5 Likelihood Approximation

Given the observations $y_{1:T}$, the likelihood of the parameter θ of the model is obtained by a *prediction error decomposition*

$$f(y_1, \ldots, y_T \mid \theta) = \prod_{t=1}^{T} f(y_t \mid y_{1:t-1}, \theta), \tag{5.24}$$

where $f(y_1 \mid y_0, \theta) = f(y_1 \mid \theta)$. We may write (5.24) as a log likelihood in the form

$$\begin{aligned} \log L(\theta) &= \log f(y_1, \ldots, y_T \mid \theta) \\ &= \sum_{t=1}^{T} \log f(y_t \mid y_{1:t-1}, \theta). \end{aligned} \tag{5.25}$$

Kitagawa (1996) suggests that by using the approximation

$$\begin{aligned} f(y_t \mid y_{1:t-1}) &= \int f(y_t \mid x_t) f(x_t \mid y_{1:t-1}) \, \mathrm{d}x_t \\ &\approx \frac{1}{N} \sum_{i=1}^{N} f\left(y_t \,\middle|\, \tilde{x}_t^{(i)}\right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \widetilde{w}_t^{(i)}, \end{aligned} \tag{5.26}$$

a pointwise approximation of the log likelihood can be obtained by

$$\log L(\theta) = \sum_{t=1}^{T} \log f(y_t \mid y_{1:t-1}) \approx \sum_{t=1}^{T} \log \left(\sum_{i=1}^{N} \widetilde{w}_t^{(i)} \right) - T \log N. \tag{5.27}$$

Therefore, the maximum likelihood estimate $\hat{\theta}$ can be estimated by maximizing the approximate log likelihood with respect to θ. However, Hürzeler and Künsch (2001) pointed out two disadvantages of this approximation. First, if we want to evaluate the likelihood at a different value of θ, then new filter and prediction samples have to be generated. On the other hand, the approximation is very noisy and some smoothing has to be applied before finding the maximum.

The likelihood approximation can also be used to evaluate the goodness of fit of several candidate models (Kitagawa, 1996), using the AIC criterion

$$\mathrm{AIC} = -2 \log L(\theta) + 2 n_p, \tag{5.28}$$

where n_p is the number of unknown parameters in the model.

Summarizing, in this Chapter we have presented state-space models and simulation-based methods known as particle filters for recursively approximating the filtering distribu-

tion. We apply these ideas in Chapter 6, where we propose state-space representations of max-stable processes and develop suitable particle filtering algorithms for these processes.

CHAPTER 6

Particle Filtering of Max-Stable Processes

In this Chapter we propose a state-space representation of M4 processes, where the state is an unobserved M4 process, and the observed process is a nonlinear transformation of the state with small additive Gaussian noise, to make the process nondegenerate. Our goal is to develop particle filter algorithms to approximate the filtering distributions of an unobserved M4 process from an observed sequence of real data.

This Chapter is organized as follows. In §6.1 we present a state-space representation of MM(1) processes, and develop three particle filtering algorithms based on the results from Chapter 4. One of the key ideas is the design of importance densities for extreme events based on local maxima. In §6.2 we provide particle filtering methods for more general MM processes. These methods are then extended for M3 and M4 processes in §6.3 and §6.4.

Throughout this Chapter we assume the parameters of the process are known. In Chapter 7 we discuss estimation of unknown parameters in the model using the same state-space representations presented here.

6.1 Particle Filtering of MM(1) Processes

In this section we concentrate on the special case of MM(1) processes, which has been discussed in detail in §4.1. We first present a state-space representation of these processes, then we develop three particle filtering algorithms based on the results from Chapter 4, and finally we present simulation results.

6.1.1 State-Space Representation

Let $\{X_t,\ t = 1, 2, \ldots\}$ be an MM(1) process as defined in (4.1), i.e.,

$$X_t = \max\{\alpha Z_t,\ (1-\alpha)Z_{t-1}\}.$$

A state-space representation of this process is motivated by the following:

1. We do not expect to observe the deterministic signature pattern of MM(1) processes in any time series data. However, we may assume that the observed process is an MM(1) process with additional measurement error.

2. It is unrealistic to expect the data to have unit Fréchet marginal distributions. However, this can be achieved after a marginal transformation based on the GEV distribution.

3. In the spirit of threshold models, we assume that the model holds only for observations above a high threshold $u > 0$. The motivation for the threshold approach is that the max-stable process is only derived as a representation for the extremes of the process, so it would not be reasonable to assume it holds for the whole process. The observations below the threshold are treated as censored. In other words, we assume that if $Y_t < u$ we have no further information about Y_t.

In summary, the observed process $\{Y_t,\ t = 1, 2, \ldots\}$ is obtained from $\{X_t\}$ after a marginal transformation for observations above a specified threshold, and the addition of measurement error. We now present a precise description of this model.

First suppose there is no observational noise. The natural marginal distribution to consider for observations above a high threshold is the GEV distribution (2.5), i.e.,

$$H_{\xi,\psi,\mu}(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\psi}\right)_+^{-1/\xi}\right\},$$

where $y_+ = \max(y, 0)$. On the other hand, it is straightforward to verify that if X is a random variable with unit Fréchet distribution function, then the transformed random

variable

$$\widetilde{X} = \mu + \psi \frac{X^{\xi} - 1}{\xi}, \tag{6.1}$$

has distribution function GEV with parameters $\xi,\ \psi,\ \mu$. The marginal distribution of the state process $\{X_t\}$ is unit Fréchet, so it follows that the relationship between $\{X_t\}$ and $\{Y_t\}$ is given by (6.1).

Now assume that the process is observed with additional measurement error. Then the state-space representation of the MM(1) process is

$$X_t = \max\{\alpha Z_t,\ (1-\alpha) Z_{t-1}\}, \qquad t \in \mathbb{N}, \tag{6.2}$$

$$Y_t = \mu + \psi \frac{X_t^{\xi} - 1}{\xi} + \epsilon_t, \qquad t \in \mathbb{N}, \tag{6.3}$$

where $\{Z_t\}$ is a sequence of iid unit Fréchet random variables, and $\{\epsilon_t\}$ is a sequence of iid normally distributed random variables with mean 0 and variance $\sigma^2 > 0$. The parameter of the MM(1) process is $\alpha \in (0, 1)$, and the parameters of the observed process are $\xi \in \mathbb{R}$, $\psi > 0$, and $\mu \in \mathbb{R}$. The vector

$$\theta = (\alpha,\ \mu,\ \log\psi,\ \xi,\ \log\sigma)' \tag{6.4}$$

denotes all the parameters in this model.

The state equation (6.2) describes the dynamics of the process $\{X_t\}$, which is a second-order Markov process by Proposition 4.1.1. The observation equation (6.3) gives the likelihood of the observation Y_t given the state variable X_t.

6.1.2 Likelihood

We assume that the model defined by (6.2) and (6.3) only holds for observations above a high threshold; the observations below the threshold need to be treated as censored observations. Following Smith, Tawn and Coles (1997, p. 252), if $Y_t < u$ the likelihood contribution is $P\{Y_t \leq u\}$, otherwise we can use the normal density. Therefore the likelihood

of the observation given the state variable is

$$f\left(y_t \mid x_t, \theta\right) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{y_t - \tilde{x}_t}{\sigma}\right), & y_t > u, \\ \Phi\left(\frac{u - \tilde{x}_t}{\sigma}\right), & y_t \leq u, \end{cases} \tag{6.5}$$

where $\tilde{x}_t$ is given by (6.1), and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution function, respectively. It should be noted that the symbols $\Phi(\cdot)$ and $\Phi_1(\cdot)$ are not related at all.

6.1.3 Basic Particle Filter

The evolution of the state process is described by the conditional distribution $F_{t+1}(x) = P\{X_{t+1} \leq x \mid \mathcal{F}_t\}$, which only depends on x_t and x_{t-1} by Proposition 4.1.1. In what follows we denote the density function of $F_{t+1}(x)$ by $f_{t+1}\left(x \mid x_{t-1:t}\right)$ or $f\left(x_{t+1} \mid x_{t-1:t}\right)$, interchangeably depending on the context.

We are interested in estimating the filtering distribution

$$f\left(x_{t+1} \mid y_{1:t+1}\right) \propto f\left(y_{t+1} \mid x_{t+1}\right) \int f\left(x_{t+1} \mid x_{t-1:t}\right) f\left(x_t \mid y_{1:t}\right). \tag{6.6}$$

The Kalman recursions and its extensions cannot be applied because the model described by (6.2) and (6.3) is nonlinear in both the state and the observation equations, and is non-Gaussian in the state. Thus our approach is to estimate (6.6) using particle filters.

We can write a basic bootstrap particle filter for MM(1) processes based on Proposition 4.1.3, which provides an algorithm to sample directly from $F_{t+1}(x)$, and the likelihood defined by (6.5). This is described in Algorithm 6.1.1.

6.1.4 Auxiliary Particle Filter

Auxiliary particle filters require a likely value of the state transition equation, such as the mean or the median, in order to generate an auxiliary variable that helps to improve the sampling. This likely value is denoted by ν_{t+1}. We write an auxiliary particle filter for MM(1) processes based on the conditional median of $F_{t+1}(x)$ derived in Corollary 4.1.5. The importance sampling step is described in Algorithm 6.1.2.

Algorithm 6.1.1 (Bootstrap Filter for MM(1) Processes).

- *Initialization*, $t = 0$.
  - For $i = 1, \ldots, N$, sample $\left(x_{-1}^{(i)}, x_0^{(i)}\right) \sim f\left(x_t, x_{t+1}\right)$.
- *Importance Sampling Step*
  1. For $i = 1, \ldots, N$, sample $x_{t+1}^{(i)} \sim f\left(x_{t+1} \,\middle|\, x_{t-1:t}^{(i)}\right)$ according to Proposition 4.1.3.
  2. For $i = 1, \ldots, N$, evaluate the importance weights $w_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right)$, where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).
- *Selection Step*
  - Resample with replacement N particles from the set $\left\{x_{t:t+1}^{(i)},\ i = 1, \ldots, N\right\}$ according to the importance weights.
  - Increase t by 1 and go to the importance sampling step.

6.1.5 Particle Filter for Local Maxima

Following Definition 3.6.1, in the case of MM(1) processes we say that y_{t+1} is a local maximum if $y_{t+1} \geq \max\{u, y_t\}$, for some large u. If y_{t+1} is a local maximum it is often the case that the importance weights in both Algorithms 6.1.1 and 6.1.2 are all very close to zero, so the filtering distribution is inadequately represented by a few or a single particle with non-zero importance weights. One reason for this behavior is that the particles are generated independently of y_{t+1} so these algorithms do not take into account that a local maximum has been observed. A second reason is that a local maximum is most likely caused by a new independent large shock, so the conditional median of the process gives little information about it. Our approach to solve this problem is to generate particles from an importance density that takes directly into account the observed y_{t+1} whenever a local maximum has been observed. In this case we do not need an auxiliary particle filter because the particles generated with this approach are close to the regions of high likelihood.

The motivation for our approach is that a local maximum represents an event in the tail of the filtering distribution. The idea is to generate particles for X_{t+1} from an importance

Algorithm 6.1.2 (Auxiliary Particle Filter for MM(1) Processes).
Importance Sampling Step

1. For $i = 1, \ldots, N$, obtain $\nu_{t+1}^{(i)}$, given by the conditional median of $F_{t+1}(x)$ derived in Corollary 4.1.5.
2. Sample an auxiliary integer variable J from the set $\{1, \ldots, N\}$ with probabilities proportional to
$$g_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(i)}, \theta\right),$$
where $f\left(y_{t+1} \,|\, \nu_{t+1}, \theta\right)$ is given by (6.5).
3. Sample $x_{t+1}^{(J)} \sim f\left(x_{t+1} \,\middle|\, x_{t-1:t}^{(J)}\right)$ according to Proposition 4.1.3.
4. Evaluate the importance weight
$$w_{t+1}^{(J)} \propto \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(J)}, \theta\right)}{f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}, \theta\right)},$$
where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).
5. Repeat steps 2–4 to produce $\left\{\left(x_{t+1}^{(i)}, w_{t+1}^{(i)}\right),\ i = 1, \ldots, N\right\}$ as required.

density that is concentrated in the tail. A natural choice is $\Phi_{1,\alpha}\left(x \mid x > u_{t+1}^*\right)$ defined in (A.10), which is the distribution of a unit Fréchet with scale parameter α given that it has exceeded some threshold u_{t+1}^* (this threshold is a function of y_{t+1} and we discuss how it is defined later). However, the support of the distribution $\Phi_{1,\alpha}\left(x \mid x > v\right)$ for some $v > 0$ does not contain the support of $F_{t+1}(x)$, so in order to apply importance sampling we define a proposal distribution of the form

$$Q_{1,\alpha}\left(x \,|\, v\right) = \gamma F_{t+1}(x) + (1-\gamma)\Phi_{1,\alpha}\left(x \mid x > v\right), \tag{6.7}$$

where $0 < \gamma < 1$. We set γ to be close to zero, e.g. $\gamma = 0.05$, so that (6.7) is concentrated on the right-tail. If we define

$$\gamma_2(v) = \gamma + \frac{1-\gamma}{1 - \exp\left\{-\alpha/v\right\}}, \tag{6.8}$$

then the corresponding density function can be written

$$q_{1,\alpha}(x \mid v) = \gamma f_{t+1}(x \mid x_{t-1:t}) + (1-\gamma)\varphi_{1,\alpha}(x \mid x > v)$$
$$= \begin{cases} \gamma\, f_{t+1}(x \mid x_{t-1:t}), & 0 \le x < v, \\ \gamma_2(v)\, \varphi_{1,\alpha}(x), & x \ge v, \end{cases} \tag{6.9}$$

where

$$\varphi_{1,\alpha}(x) = \frac{\alpha}{x^2} \exp\left\{-\frac{\alpha}{x}\right\}, \qquad x > 0. \tag{6.10}$$

We now describe how to define a suitable threshold u^*_{t+1}. If y_{t+1} is a local maximum then the noise term in the observation equation is small compared to the unobserved state variable X_{t+1}, so we can approximate

$$X^*_{t+1} \approx \left(1 + \xi \frac{y_{t+1} - \mu}{\psi}\right)^{1/\xi}_{+}. \tag{6.11}$$

Now assume that X^*_{t+1} is the median of a distribution $\Phi_{1,\alpha}\left(x \mid x > v^*_{t+1}\right)$ defined by (A.10) for some threshold v^*_{t+1}. Then by (A.15) we have

$$X^*_{t+1} \approx 2v^*_{t+1} + \frac{\alpha}{2}, \tag{6.12}$$

from which we obtain

$$v^*_{t+1} \approx \frac{1}{2}\left(X^*_{t+1} - \frac{\alpha}{2}\right). \tag{6.13}$$

Hence, we define

$$u^*_{t+1} = \max\left\{v^*_{t+1}, \frac{1-\alpha}{\alpha} x_t\right\}. \tag{6.14}$$

This definition ensures that the importance weights are easy to compute.

In summary, if y_{t+1} is a local maximum then our approach is to generate particles for $x^{(i)}_{t+1}$, $i = 1, \ldots, N$ from the distribution $Q_{1,\alpha}\left(x \,\middle|\, u^{*(i)}_{t+1}\right)$. The importance weights for this algorithm are evaluated as follows. From Proposition 4.1.1 it is straightforward to verify that

$$f_{t+1}(x \mid x_{t-1:t}) = \varphi_{1,\alpha}(x) \qquad x \ge \frac{1-\alpha}{\alpha} x_t. \tag{6.15}$$

Therefore the importance weights are

$$w_{t+1}^{(i)} \propto \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right) f\left(x_{t+1}^{(i)} \,\middle|\, x_{t-1:t}^{(i)}\right)}{q_{1,\alpha}\left(x_{t+1}^{(i)} \,\middle|\, u_{t+1}^{*(i)}\right)}$$
$$= \begin{cases} f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right) \Big/ \gamma, & 0 \le x_{t+1}^{(i)} < u_{t+1}^{*(i)}, \\ f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right) \Big/ \gamma_2(u_{t+1}^{*(i)}), & x_{t+1}^{(i)} \ge u_{t+1}^{*(i)}. \end{cases} \tag{6.16}$$

The importance sampling step for local maxima is summarized in Algorithm 6.1.3.

Algorithm 6.1.3 (Particle Filter for Local Maxima of MM(1) Processes).
Importance Sampling Step
For $i = 1, \ldots, N$,

1. Calculate $u_{t+1}^{*(i)}$ according to (6.14).
2. Generate $x_{t+1}^{(i)} \sim Q_{1,\alpha}\left(x \,\middle|\, u_{t+1}^{*(i)}\right)$, where $Q_{1,\alpha}(x \mid v)$ is defined by (6.7).
3. Evaluate the importance weights $w_{t+1}^{(i)}$ according to (6.16).

6.1.6 Simulation Results

We compare the particle filters for MM(1) processes with simulated data. In what follows we refer to Algorithm 6.1.1 as the basic particle filter, and we refer to Algorithm 6.1.3 as the improved particle filter for local maxima.

We simulate a sample path of size $T = 1000$ of an MM(1) process and the corresponding observed process. We use $\alpha = 0.8$, $\mu = 0$, $\psi = 0.5$, $\xi = 0.18$, and $\sigma = 0.2$. We obtain filtering distributions from each method using a threshold u set at the 0.97 quantile of the observed process, and $N = 2000$ particles.

Figure 6.1 shows part of the observed exceedances over the threshold (top); and the corresponding filtered process in log scale (solid line) from the particle filter for local maxima, including 95% posterior bands (dotted), and the actual simulated process (dashed) for reference. The filtered process follows closely the actual process whenever there is an exceedance above the threshold. On the other hand, the median of the filtered process for observations below the threshold is expected to fluctuate around $1/\log 2$ (the median of

the unit Fréchet distribution function) because the only information available for censored observations is that they are below the threshold. However, the confidence bands are wide enough to accommodate the fluctuations of the actual process.

In order to compare the different methods, we concentrate on the filtered process corresponding to exceedances over the threshold. It is expected that the filtered process above the threshold agrees with the actual process, so a scatter plot of the filtered against the actual process should lie on the 45 degree reference line. Figure 6.2 shows this plot for both the basic particle filter (left) and the improved particle filter (right). The line segments show pointwise 95% intervals of the posterior distribution of the filters.

In some of the confidence intervals from the basic particle filter, especially for very large observations, the endpoints of the intervals coincide with the median—an indication of a degenerate distribution. On the other hand, there are no degenerate confidence intervals from the improved particle filter, and all of these intervals intersect the 45 degree reference line. These results support that the particle filter for local maxima is doing a better job at approximating the filtering distribution.

In summary, we propose Algorithm 6.1.3 for filtering local maxima of MM(1) processes, and either Algorithm 6.1.1 or Algorithm 6.1.2 for the rest of the series. The results from our simulation experiments suggest that there are no gains from using the auxiliary particle filter over the basic particle filter because these are mainly used for censored observations. The same behavior will be observed for more general processes in the following sections.

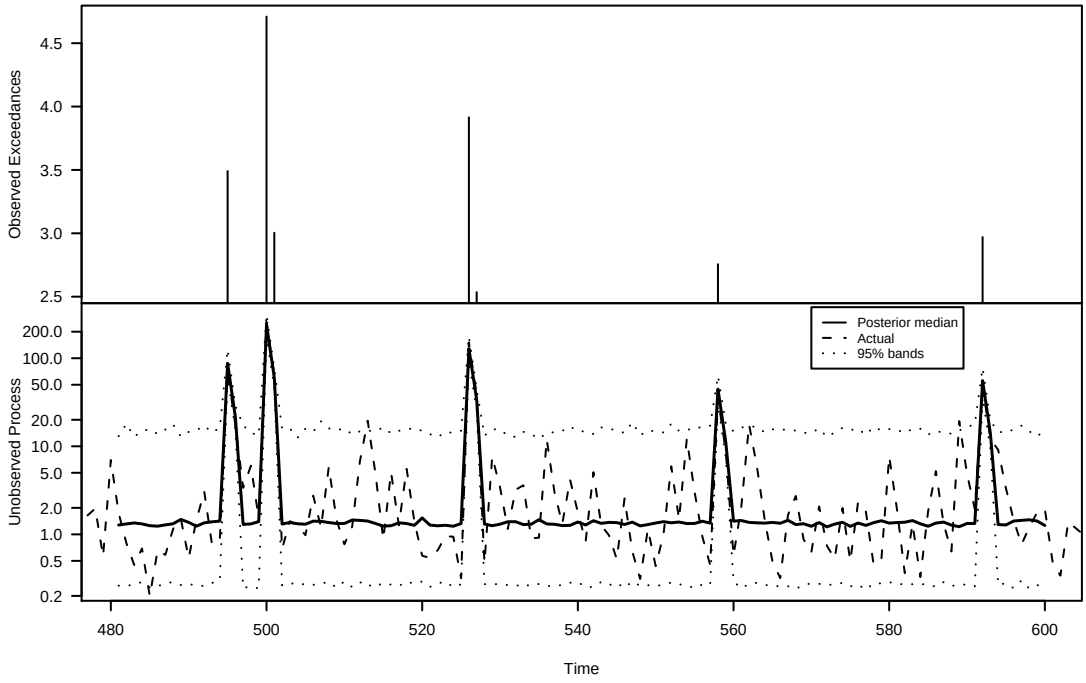

Figure 6.1: *Observed Exceedances and Filtered MM(1) Process.*

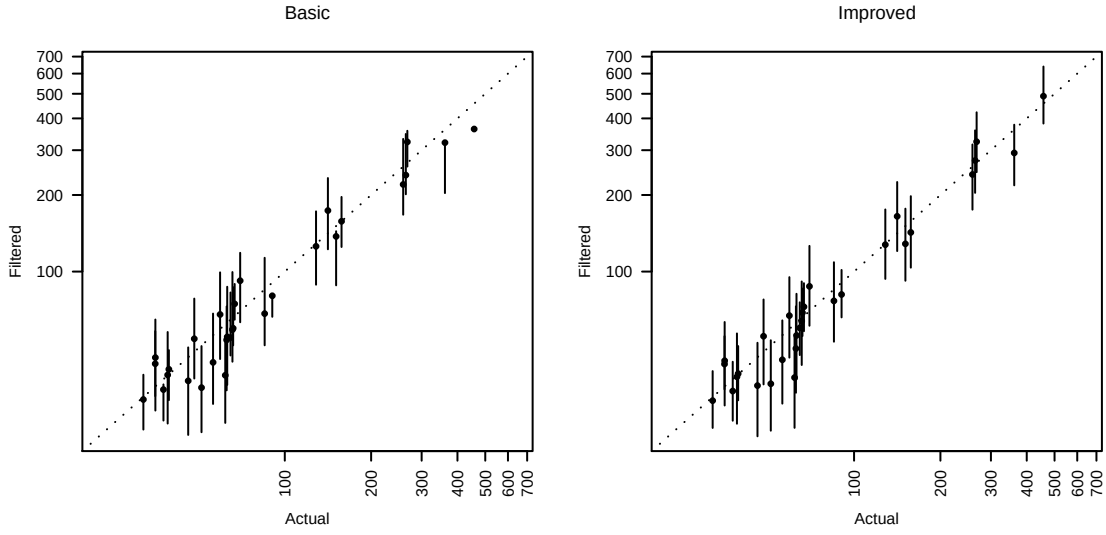

Figure 6.2: *Scatter Plots of Filtered vs. Actual MM(1) Process. Basic particle filter (left) and improved particle filter for local maxima (right).*

6.2 Particle Filtering of MM Processes

In this section we propose particle filtering algorithms for MM processes. The methods are developed separately from the MM(1) case because of the lack of explicit expressions for the conditional distributions of general MM processes.

As in §6.1, we first present the state-space representation, then we develop particle filtering algorithms, and finally we present simulation results.

6.2.1 State-Space Representation

Following the developments in §6.1, the state-space representation of MM processes is

$$X_t = \max_{0 \le k \le K} \alpha_k Z_{t-k}, \qquad t \in \mathbb{N}, \tag{6.17}$$

$$Y_t = \mu + \psi \frac{X_t^{\xi} - 1}{\xi} + \epsilon_t, \qquad t \in \mathbb{N}, \tag{6.18}$$

where $\{Z_t\}$ is a sequence of iid unit Fréchet random variables, and $\{\epsilon_t\}$ is a sequence of iid normally distributed random variables with mean 0 and variance $\sigma^2 > 0$. The vector

$$\theta = (\alpha,\ \mu,\ \log\psi,\ \xi,\ \log\sigma)' \tag{6.19}$$

denotes all the parameters in this model, where $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_K)'$ satisfies $\alpha_k \ge 0$ and $\sum_k \alpha_k = 1$, $\xi \in \mathbb{R}$, $\psi > 0$, and $\mu \in \mathbb{R}$.

Remark. To simplify the exposition we only consider models where $\alpha_k = 0$ for $k < 0$. However, it is straightforward to generalize the algorithms for models with forward lags.

6.2.2 Basic Particle Filter

The conditional distribution of a general MM process is not known explicitly. However, we can describe the evolution of the state in terms of the independent shocks $\{Z_t\}$. The idea is that given the sequence $\{Z_t\}$, the process $\{X_t\}$ is completely determined.

The prediction step in this case is to obtain the distribution of X_{t+1} given $Z_{t-K+1:t}$, which only requires sampling a new innovation Z_{t+1} and then applying (6.17). The advantage of using this approach is that the extension for M4 processes is straightforward. The

likelihood of MM processes is still given by (6.5). A basic bootstrap particle filter for MM processes is described in Algorithm 6.2.1.

Algorithm 6.2.1 (Bootstrap Filter for MM Processes).

- *Initialization*, $t = 0$.
  - For $i = 1, \ldots, N$, sample $\left(Z_{-K+1}^{(i)}, Z_{-K+2}^{(i)}, \ldots, Z_0^{(i)}\right) \sim \Phi_1(x)$.
- *Importance Sampling Step*
  1. For $i = 1, \ldots, N$, sample $Z_{t+1}^{(i)} \sim \Phi_1(x)$ and combine with $Z_{t-K+1:t}^{(i)}$ to obtain $x_{t+1}^{(i)}$ from (6.17).
  2. For $i = 1, \ldots, N$, evaluate the importance weights $w_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right)$, where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).
- *Selection Step*
  - Resample with replacement N particles from the set $\left\{Z_{t-K+2:t+1}^{(i)},\ i = 1, \ldots, N\right\}$ according to the importance weights.
  - Increase t by 1 and go to the importance sampling step.

6.2.3 Auxiliary Particle Filter

An auxiliary particle filter for MM processes is based on Proposition 6.2.2, which shows that MM processes have an upper bound given by a max-autoregressive process. This result can be applied to obtain likely values for the auxiliary particle filter.

Proposition 6.2.2. *Suppose* $\{X_t\}$ *is an MM process as defined in* (6.17). *Then*

$$X_{t+1} \leq \max\left\{\alpha_0 Z_{t+1},\, \beta X_t\right\}, \tag{6.20}$$

where

$$\beta = \max_{0 \leq k < K} \frac{\alpha_{k+1}}{\alpha_k}. \tag{6.21}$$

Proof. By definition of the process we have $Z_{t-k} \leq X_t / \alpha_k$ for each $k = 0, 1, \ldots, K$. Then

$$X_{t+1} = \max\left\{\alpha_0 Z_{t+1},\, \alpha_1 Z_t,\, \ldots, \alpha_K Z_{t-K+1}\right\}$$

$$
\begin{aligned}
&\leq \max\left\{\alpha_0 Z_{t+1},\, \frac{\alpha_1}{\alpha_0}X_t,\, \ldots,\, \frac{\alpha_K}{\alpha_{K-1}}X_t\right\} \\
&= \max\left\{\alpha_0 Z_{t+1},\, \max_{0\leq k<K}\frac{\alpha_{k+1}}{\alpha_k}X_t\right\} \\
&= \max\left\{\alpha_0 Z_{t+1},\, \beta X_t\right\}, \qquad (6.22)
\end{aligned}
$$

where β is given by (6.21). □

We obtain likely values for the auxiliary particle filter as follows. Given X_t, generate Z^* from a unit Fréchet distribution and define

$$
\nu_{t+1} = \max\left\{\alpha_0 Z^*,\, \beta X_t\right\}. \qquad (6.23)
$$

Alternatively, since the median of a unit Fréchet is $1/\log 2$, we can define

$$
\nu_{t+1} = \max\left\{\frac{\alpha_0}{\log 2},\, \beta X_t\right\}. \qquad (6.24)
$$

The importance sampling step of the auxiliary particle filter for MM processes is described in Algorithm 6.2.3. This algorithm suffers from the same drawback than the corresponding method for MM(1) process. If a local maximum is observed then often the importance weights are all zero.

6.2.4 Particle Filter for Local Maxima

In this section we discuss a method to improve the auxiliary particle filter for observations that are identified as local maxima according to Definition 3.6.1. We follow closely §6.1.5, where we develop particle filters for local maxima of MM(1) processes.

If y_{t+1} is a local maximum then it represents a tail event of the filtering distribution. If we do not take into account y_{t+1} to draw the proposals for the filtering distribution then this tail event is poorly represented by the particles. Hence, we have to move the particles close to y_{t+1}, so that the likelihood of y_{t+1} given the particles is not negligible.

Our importance sampler is based on a proposal distribution of the form

$$
Q\left(z \mid v\right) = \gamma\Phi_1(z) + (1-\gamma)\Phi_{1,1}\left(z \mid z > v\right), \qquad (6.25)
$$

Algorithm 6.2.3 (Auxiliary Particle Filter for MM Processes).
Importance Sampling Step

1. For $i = 1, \ldots, N$, obtain $\nu_{t+1}^{(i)}$ from (6.23) or (6.24).
2. Sample an auxiliary integer variable J from the set $\{1, \ldots, N\}$ with probabilities proportional to
$$g_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(i)}, \theta\right),$$
where $f\left(y_{t+1} \,|\, \nu_{t+1}, \theta\right)$ is given by (6.5).
3. Sample $Z_{t+1} \sim \Phi_1(x)$ and combine with $Z_{t-K+1:t}^{(J)}$ to obtain $x_{t+1}^{(J)}$ from (6.17).
4. Evaluate the importance weight
$$w_{t+1}^{(J)} \propto \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(J)}, \theta\right)}{f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}, \theta\right)},$$
where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).
5. Repeat steps 2–4 to produce $\left\{\left(x_{t+1}^{(i)}, w_{t+1}^{(i)}\right),\ i = 1, \ldots, N\right\}$ as required.

where $0 < \gamma < 1$, $\Phi_1(z) = \exp\{-1/z\}$, $z > 0$, and $\Phi_{1,1}\left(z \mid z > v\right)$ is defined in (A.10), i.e., the distribution of a unit Fréchet random variable given it has exceeded some threshold v. We set γ to be close to zero, e.g. $\gamma = 0.05$, so that (6.25) is concentrated on the right-tail. If we define

$$h(z, v, \gamma) = \gamma + \frac{1 - \gamma}{1 - \exp\{-1/v\}} \mathbb{I}\left(z \geq v\right). \tag{6.26}$$

then the corresponding density function can be written

$$\begin{aligned} q\left(z \mid v\right) &= \gamma\varphi_1(z) + (1 - \gamma)\varphi_1\left(z \mid z > v\right) \\ &= \varphi_1(z)h(z, v, \gamma). \end{aligned} \tag{6.27}$$

where $\varphi_1(z) = z^{-2}\exp\{-1/z\}$, $z > 0$, is the unit Fréchet density.

We now construct a suitable threshold, denoted u_{t+1}^*, for the proposal distribution (6.25). If y_{t+1} is a local maximum then the noise term in the observation equation (6.18)

is small compared to the state variable X_{t+1}, so we can approximate

$$X^*_{t+1} \approx \left(1 + \xi \frac{y_{t+1} - \mu}{\psi}\right)^{1/\xi}_{+}. \tag{6.28}$$

Furthermore, if y_{t+1} is a local maximum then we assume that X^*_{t+1} comes from a new large shock Z^*_{t+1}, so that

$$X^*_{t+1} = \alpha_0 Z^*_{t+1} > \max_{1 \le k \le K} \alpha_k Z_{t-k}, \tag{6.29}$$

which gives

$$Z^*_{t+1} = \frac{X^*_{t+1}}{\alpha_0}. \tag{6.30}$$

The value Z^*_{t+1} gives a rough estimate of the location of the actual shock that generated the local maximum. Hence, our approach is to assume that Z^*_{t+1} is the median of the distribution of a unit Fréchet random variable given it has exceeded the threshold u^*_{t+1}. Then by (A.15) we have

$$Z^*_{t+1} \approx 2u^*_{t+1} + \frac{1}{2}, \tag{6.31}$$

from which we obtain

$$u^*_{t+1} \approx \frac{1}{2}\left(Z^*_{t+1} - \frac{1}{2}\right). \tag{6.32}$$

Note that $h(\gamma, z, v) > 0$, and $u^*_{t+1} > 0$ is a constant for all the particles. The importance weights for this algorithm are

$$\begin{aligned} w^{(i)}_{t+1} &\propto \frac{f\left(y_{t+1} \,\middle|\, x^{(i)}_{t+1}, \theta\right) \varphi_1\left(z^{(i)}_{t+1}\right)}{q\left(z^{(i)}_{t+1} \,\middle|\, u^*_{t+1}\right)} \\ &= \frac{f\left(y_{t+1} \,\middle|\, x^{(i)}_{t+1}, \theta\right)}{h\left(z^{(i)}_{t+1}, u^*_{t+1}, \gamma\right)}. \end{aligned} \tag{6.33}$$

We summarize the importance sampling step of the algorithm for local maxima in Algorithm 6.2.4. This algorithm generates particles for extreme events (local maxima) that are closer to the current observation. The advantage is that the importance weights have less variability. It is important to note that Algorithm 6.2.4 only applies to observations that are identified as local maxima. If an observation is *not* a local maximum then we can

use Algorithm 6.2.1 or Algorithm 6.2.3.

Algorithm 6.2.4 (Particle Filter for Local Maxima of MM Processes).
Importance Sampling Step

1. Calculate u^*_{t+1} according to (6.32).
2. For $i = 1, \ldots, N$, generate $x^{(i)}_{t+1}$ as follows.
   - Generate $Z^{(i)}_{t+1} \sim Q\left(z \mid u^*_{t+1}\right)$, where $Q(z \mid v)$ is defined by (6.25).
   - Combine $Z^{(i)}_{t+1}$ with $Z^{(i)}_{t-K+1:t}$ to obtain $x^{(i)}_{t+1}$ from the state equation (6.17).
3. For $i = 1, \ldots, N$, evaluate the importance weights $w^{(i)}_{t+1}$ according to (6.33).

6.2.5 Simulation Results

We compare the particle filters for MM processes in the same way as for MM(1) processes. In this case we compare Algorithm 6.2.1 (basic) and Algorithm 6.2.4 (improved).

We simulate a sample path of size $T = 1000$ of an MM(2) process with parameter $\alpha = (0.6, 0.3, 0.1)$, and the corresponding observed process with $\mu = 0$, $\psi = 0.5$, $\xi = 0.18$, and $\sigma = 0.2$. We obtain filtering distributions from each method using a threshold u set at the 0.97 quantile of the observed process, and $N = 2000$ particles.

Figure 6.3 shows part of the observed exceedances over the threshold (top); and the corresponding filtered process in log scale (solid line) from the particle filter for local maxima, including 95% posterior bands (dotted), and the actual simulated process (dashed) for reference.

Figure 6.4 shows scatter plots of the filtered against the actual process for both the basic particle filter (left) and the improved particle filter (right). The improved particle filter performs better as expected.

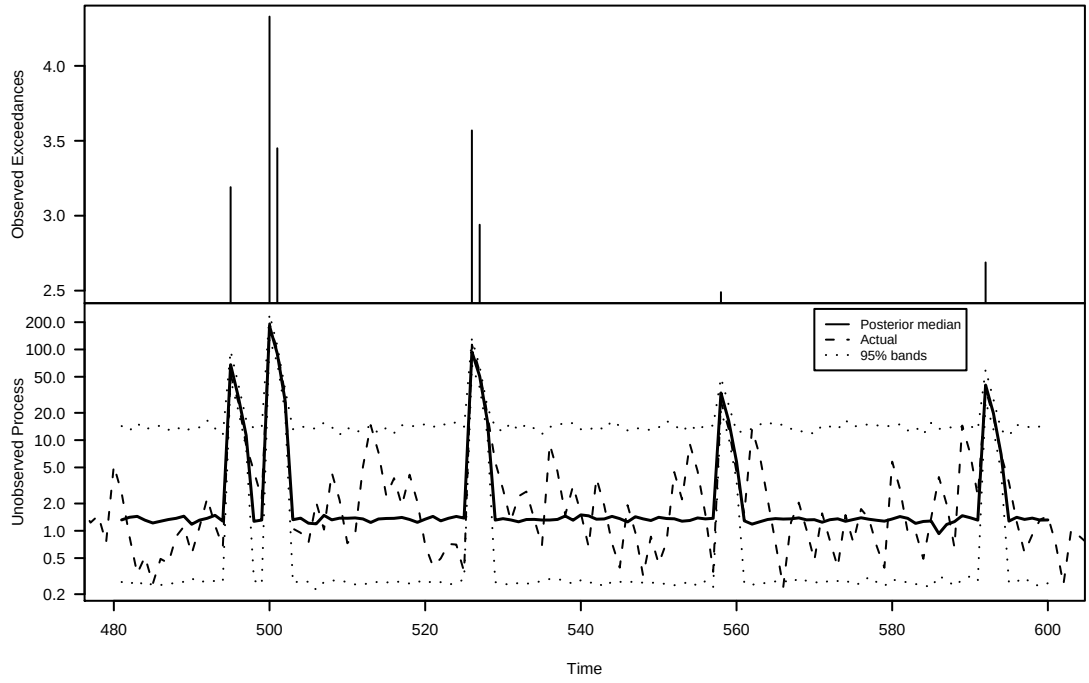

Figure 6.3: *Observed Exceedances and Filtered MM(2) Process.*

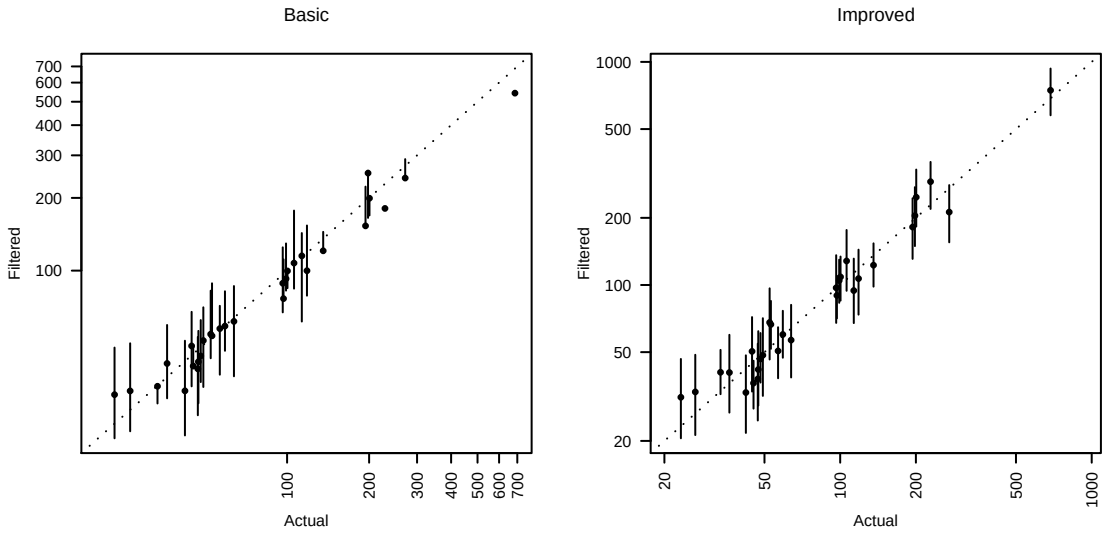

Figure 6.4: *Scatter Plots of Filtered vs. Actual MM(2) Process. Basic particle filter (left) and improved particle filter for local maxima (right).*

6.3 Particle Filtering of M3 Processes

In this section we develop particle filtering methods for M3 processes by extending the methods for MM processes.

6.3.1 State-Space Representation

The state-space representation of the M3 process is

$$X_t = \max_{1\le\ell\le L} \max_{0\le k\le K} \alpha_{\ell k} Z_{\ell,t-k}, \qquad t \in \mathbb{N}, \tag{6.34}$$

$$Y_t = \mu + \psi \frac{X_t^{\xi} - 1}{\xi} + \epsilon_t, \qquad t \in \mathbb{N}, \tag{6.35}$$

where $\{Z_{\ell t}\}$ is a sequence of iid unit Fréchet random variables, and $\{\epsilon_t\}$ is a sequence of iid normally distributed random variables with mean 0 and variance $\sigma^2 > 0$. The vector

$$\theta = (\alpha,\ \mu,\ \log\psi,\ \xi,\ \log\sigma)', \tag{6.36}$$

denotes all the parameters in this model, where $\alpha = \{\alpha_{\ell k}\}$ is a sequence of nonnegative constants satisfying $\sum_\ell \sum_k \alpha_{\ell k} = 1$, $\xi \in \mathbb{R}$, $\psi > 0$, and $\mu \in \mathbb{R}$.

Remark. To simplify the exposition we only consider models where $\alpha_{\ell k} = 0$ for $k < 0$. However, it is straightforward to generalize the algorithms for models with forward lags.

6.3.2 Basic Particle Filter

A straightforward extension of Algorithm 6.2.1 gives the basic particle filter algorithm for M3 processes. In this case we have L signature patterns, so at each time t prediction of the new state requires generating L new shocks $Z_{\ell,t+1}$, $\ell = 1, \ldots, L$.

To initialize the algorithm generate $\left\{Z_{\ell,j}^{(i)}\right\}$ for $1 \le \ell \le L$, $-K+1 \le j \le 0$, and $i = 1, \ldots, N$. The importance sampling step is described in Algorithm 6.3.1.

6.3.3 Auxiliary Particle Filter

In this section we derive an extension of the inequality derived in Proposition 6.2.2, and present an auxiliary particle filter for M3 processes.

Algorithm 6.3.1 (Bootstrap Filter for M3 Processes).
Importance Sampling Step

1. For $i = 1, \ldots, N$, sample $Z_{\ell,t+1} \sim \Phi_1(x)$, $\ell = 1, \ldots, L$ and combine with $Z_{\ell,t}^{(i)}$, $Z_{\ell,t-1}^{(i)}, \ldots, Z_{\ell,t-K+1}^{(i)}$, $\ell = 1, \ldots, L$ to obtain $x_{t+1}^{(i)}$ from (6.34).
2. For $i = 1, \ldots, N$, evaluate the importance weights $w_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, x_{t+1}^{(i)}, \theta\right)$, where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).

Proposition 6.3.2. *Suppose $\{X_t\}$ is an M3 process as defined in (6.34). Then*

$$X_{t+1} \leq \max_{1 \leq \ell \leq L} \left\{\alpha_{\ell,0} Z_{\ell,t+1},\, \beta X_t\right\}, \tag{6.37}$$

where

$$\beta = \max_{1 \leq \ell \leq L} \max_{0 \leq k < K} \frac{\alpha_{\ell,k+1}}{\alpha_{\ell k}}. \tag{6.38}$$

Proof. By definition of the process we have $Z_{\ell,t-k} \leq X_t / \alpha_{\ell k}$ for each pair (ℓ, k). Then

$$\begin{aligned} X_{t+1} &= \max_{\ell} \left\{\alpha_{\ell,0} Z_{\ell,t+1},\, \alpha_{\ell,1} Z_{\ell,t},\, \ldots, \alpha_{\ell,K} Z_{\ell,t-K+1}\right\} \\ &\leq \max_{\ell} \left\{\alpha_{\ell,0} Z_{\ell,t+1},\, \frac{\alpha_{\ell,1}}{\alpha_{\ell,0}} X_t,\, \ldots, \frac{\alpha_{\ell,K}}{\alpha_{\ell,K-1}} X_t\right\} \\ &= \max_{\ell} \left\{\alpha_{\ell,0} Z_{\ell,t+1},\, \max_{\ell} \max_{0 \leq k < K} \frac{\alpha_{\ell,k+1}}{\alpha_{\ell k}} X_t\right\} \\ &= \max_{\ell} \left\{\alpha_{\ell,0} Z_{\ell,t+1},\, \beta X_t\right\}, \end{aligned} \tag{6.39}$$

where β is given by (6.38). □

We obtain likely values for the auxiliary particle filter as follows. Given X_t, generate iid innovations $Z_1^*, \ldots, Z_L^*$ from a unit Fréchet distribution and then set

$$\nu_{t+1} = \max_{\ell} \left\{\alpha_{\ell,0} Z_{\ell}^*,\, \beta X_t\right\}. \tag{6.40}$$

Alternatively, we can obtain a likely value for the auxiliary particle filter by taking the

maximum of the median from each innovation and the current state, i.e.,

$$\nu_{t+1} = \max_{\ell} \left\{ \frac{\alpha_{\ell,0}}{\log 2}, \beta X_t \right\}. \tag{6.41}$$

The importance sampling step of the auxiliary particle filter for M3 processes is described in Algorithm 6.3.3.

Algorithm 6.3.3 (Auxiliary Particle Filter for M3 Processes).
Importance Sampling Step

1. For $i = 1, \ldots, N$, obtain $\nu_{t+1}^{(i)}$ from (6.40) or (6.41).
2. Sample an auxiliary integer variable J from the set $\{1, \ldots, N\}$ with probabilities proportional to
$$g_{t+1}^{(i)} \propto f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(i)}, \theta\right),$$
where $f\left(y_{t+1} \,|\, \nu_{t+1}, \theta\right)$ is given by (6.5).
3. Sample $Z_{\ell,t+1} \sim \Phi_1(x)$, $\ell = 1, \ldots, L$ and combine with $Z_{\ell,t}^{(J)}, Z_{\ell,t-1}^{(J)}, \ldots, Z_{\ell,t-K+1}^{(J)}$, $\ell = 1, \ldots, L$ to obtain $x_{t+1}^{(J)}$ from (6.34).
4. Evaluate the importance weight
$$w_{t+1}^{(J)} \propto \frac{f\left(y_{t+1} \,\middle|\, x_{t+1}^{(J)}, \theta\right)}{f\left(y_{t+1} \,\middle|\, \nu_{t+1}^{(J)}, \theta\right)},$$
where $f\left(y_{t+1} \,|\, x_{t+1}, \theta\right)$ is given by (6.5).
5. Repeat steps 2–4 to produce $\left\{\left(x_{t+1}^{(i)}, w_{t+1}^{(i)}\right),\ i = 1, \ldots, N\right\}$ as required.

6.3.4 Particle Filter for Local Maxima

We propose an algorithm based on local maxima to improve the sampling. The difference from MM processes is that M3 processes have L signature patterns, which means that a local maximum may arise from a single large shock in one of the L independent innovation sequences.

In analogy with Algorithm 6.2.4, we make use of the importance density $Q\left(z \,|\, u_{t+1}^*\right)$ defined by (6.25). In this case u_{t+1}^* is calculated as follows. First obtain X_{t+1}^* from (6.28),

then define

$$Z^*_{t+1} = \frac{X^*_{t+1}}{\max_\ell \max_k \alpha_{\ell k}}, \tag{6.42}$$

and then obtain u^*_{t+1} from (6.32). [1]

If a local maximum is observed then the algorithm we propose is to randomly sample an index ℓ^* from the set $\{1, \ldots, L\}$, where each ℓ can be chosen with probability $\alpha^+_\ell = \sum_k \alpha_{\ell k}$, which is the relative frequency of the ℓth signature pattern. Then given ℓ^*, sample a large candidate shock $Z_{\ell^*,t+1} \sim Q\left(z \mid u^*_{t+1}\right)$, where $Q\left(z \mid v\right)$ is defined by (6.25). Finally, sample the remaining $Z_{\ell,t+1}$, $\ell \neq \ell^*$ from a unit Fréchet distribution.

The effect of randomly generating ℓ^* is that the importance density is of the form

$$\begin{aligned} q\left(z_\ell,\ 1 \le \ell \le L \mid u\right) &= \sum_\ell \alpha^+_\ell \, q\left(z_\ell \mid u\right) \prod_{j \neq \ell} \varphi_1(z_j) \\ &= \sum_\ell \alpha^+_\ell h(z_\ell, u, \gamma) \varphi_1(z_\ell) \prod_{j \neq \ell} \varphi_1(z_j) \\ &= \sum_\ell \alpha^+_\ell h(z_\ell, u, \gamma) \prod_{j=1}^{L} \varphi_1(z_j) \\ &= \prod_{j=1}^{L} \varphi_1(z_j) \sum_\ell \alpha^+_\ell h(z_\ell, u, \gamma), \end{aligned} \tag{6.43}$$

where $h(z, u, \gamma)$ is defined by (6.26). Hence the importance weights are

$$\begin{aligned} w^{(i)}_{t+1} &\propto \frac{f\left(y_{t+1} \,\middle|\, x^{(i)}_{t+1}, \theta\right) \prod_{\ell=1}^{L} \varphi_1\left(z^{(i)}_{\ell,t+1}\right)}{q\left(z^{(i)}_{\ell,t+1},\ 1 \le \ell \le L \,\middle|\, u^*_{t+1}\right)} \\ &= \frac{f\left(y_{t+1} \,\middle|\, x^{(i)}_{t+1}, \theta\right)}{\sum_\ell \alpha^+_\ell h\left(z^{(i)}_{\ell,t+1}, u^*_{t+1}, \gamma\right)}. \end{aligned} \tag{6.44}$$

The details of the importance sampling step of this algorithm are given in Algorithm 6.3.4.

[1] Note that there is nothing special about (6.42) – we only need a rough estimate of the shock inducing the local maximum, so alternative definitions may be used. For example, if we assume that the shock comes from a specific ℓ, say ℓ^*, then an alternative definition for the denominator of (6.32) is $\max_k \alpha_{\ell^*,k}$. However, we do not consider this.

Algorithm 6.3.4 (Particle Filter for Local Maxima of M3 Processes).
Importance Sampling Step

1. Calculate u^*_{t+1} according to (6.42) and (6.32).
2. For $i = 1, \ldots, N$, generate $x^{(i)}_{t+1}$ as follows.
   - Generate an index $\ell^{(i)^*}$ from the set $\{1, \ldots, L\}$ with probability $\alpha^+_\ell = \sum_k \alpha_{\ell k}$.
   - Generate $Z^{(i)^*}_{\ell^{(i)^*},t+1} \sim Q\left(z \mid u^*_{t+1}\right)$, where $Q(z \mid v)$ is defined by (6.25).
   - Generate $Z^{(i)^*}_{\ell,t+1} \sim \Phi_1(x)$, for $\ell \neq \ell^{(i)^*}$.
   - Combine $Z^{(i)}_{\ell,t+1},\ \ell = 1, \ldots, L$ with $Z^{(i)}_{\ell,t}, Z^{(i)}_{\ell,t-1}, \ldots, Z^{(i)}_{\ell,t-K+1}$ to obtain $x^{(i)}_{t+1}$ from the state equation (6.34).
3. For $i = 1, \ldots, N$, evaluate the importance weights $w^{(i)}_{t+1}$ according to (6.44).

6.4 Particle Filtering of M4 Processes

In this section we present two particle filtering algorithms for M4 processes based on extensions of the algorithms developed in §6.3. The major differences from the univariate case are the evaluation of the likelihood, the definition of local maxima, and the relative frequency of the signature patterns.

6.4.1 State-Space Representation

The state-space representation of the M4 process is

$$X_{td} = \max_\ell \max_k \alpha_{\ell k d} Z_{\ell,t-k}, \qquad t \in \mathbb{N}, \quad 1 \le d \le D, \tag{6.45}$$

$$Y_{td} = \mu_d + \psi_d \frac{X^{\xi_d}_{td} - 1}{\xi_d} + \epsilon_{td}, \qquad t \in \mathbb{N}, \quad 1 \le d \le D, \tag{6.46}$$

where $\{Z_{\ell t}\}$ is a sequence of iid unit Fréchet random variables, and $\{\epsilon_{td}\}$ is a sequence of iid normally distributed random variables with mean 0 and variance $\sigma^2_d > 0$ for each $d = 1, \ldots, D$. The vector θ is used to denote all of the parameters in this model, where $\alpha = \{\alpha_{\ell k d}\}$ is a sequence of nonnegative constants satisfying

$$\sum_\ell \sum_k \alpha_{\ell k d} = 1, \qquad \text{for each } d = 1, \ldots, D, \tag{6.47}$$

and $\xi_d \in \mathbb{R}$, $\psi_d > 0$, and $\mu_d \in \mathbb{R}$ for each $d = 1, \ldots, D$.

We assume that for fixed t, $\{\epsilon_{td}\}$ are independent for each d, so the likelihood is the product of the individual components, i.e.,

$$f\left(\boldsymbol{y}_t \,|\, \boldsymbol{x}_t, \theta\right) = \prod_{d=1}^{D} f\left(y_{td} \,|\, x_{td}, \theta\right), \tag{6.48}$$

where $f\left(y_{td} \,|\, x_{td}, \theta\right)$ is given by (6.5).

6.4.2 Basic Particle Filter

To initialize the algorithm generate $\left\{Z_{\ell,j}^{(i)}\right\}$ for $1 \leq \ell \leq L$, $-K+1 \leq j \leq 0$, and $i = 1, \ldots, N$. The importance sampling step is described in Algorithm 6.4.1.

Algorithm 6.4.1 (Bootstrap Filter for M4 Processes).
Importance Sampling Step

1. For $i = 1, \ldots, N$, sample $Z_{\ell,t+1} \sim \Phi_1(x)$, $\ell = 1, \ldots, L$ and combine with $Z_{\ell,t}^{(i)}$, $Z_{\ell,t-1}^{(i)}, \ldots, Z_{\ell,t-K+1}^{(i)}$, $\ell = 1, \ldots, L$ to obtain $x_{t+1,d}^{(i)}$, $d = 1, \ldots, D$ from (6.45).
2. For $i = 1, \ldots, N$, evaluate the importance weights $w_{t+1}^{(i)}$ according to (6.48).

6.4.3 Particle Filter for Local Maxima

From Definition 3.6.5, we say that y_{td} is a local maximum if $y_{td} \geq \max\left\{u, \boldsymbol{y}_{t-K:t+K}\right\}$. If a local maximum is observed then we proceed as in Algorithm 6.3.4, however in this case the relative frequency of the ℓth signature pattern is proportional to $\alpha_\ell^\ddagger = \sum_k \max_d \alpha_{\ell k d}$.

A suitable threshold u_{t+1}^* for the proposal distribution (6.25) is calculated as follows. Suppose the local maximum is $y_{t+1,d}$ then we approximate

$$X_{t+1,d}^* \approx \left(1 + \xi_d \frac{y_{t+1,d} - \mu_d}{\psi_d}\right)_+^{1/\xi_d}, \tag{6.49}$$

then define

$$Z_{t+1}^* = \frac{X_{t+1,d}^*}{\max_\ell \max_k \alpha_{\ell k d}}, \tag{6.50}$$

and then obtain u_{t+1}^* from (6.32).

The importance weights are

$$w_{t+1}^{(i)} \propto \frac{f\left(\boldsymbol{y}_{t+1} \mid \boldsymbol{x}_{t+1}^{(i)}, \theta\right) \prod_{\ell=1}^{L} \varphi_1\left(z_{\ell,t+1}^{(i)}\right)}{q\left(z_{\ell,t+1}^{(i)},\ 1 \le \ell \le L \,\middle|\, u_{t+1}^{*}\right)}$$
$$= \frac{f\left(\boldsymbol{y}_{t+1} \mid \boldsymbol{x}_{t+1}^{(i)}, \theta\right)}{\sum_{\ell} \alpha_{\ell}^{\ddagger} h\left(z_{\ell,t+1}^{(i)}, u_{t+1}^{*}, \gamma\right)}, \tag{6.51}$$

where $h(z, u, \gamma)$ is defined by (6.26).

The importance sampling step of this algorithm is given in Algorithm 6.4.2.

Algorithm 6.4.2 (Particle Filter for Local Maxima of M4 Processes).
Importance Sampling Step

1. Calculate u_{t+1}^{*} according to (6.50) and (6.32).
2. For $i = 1, \ldots, N$, generate $x_{t+1,d}^{(i)}$, $d = 1, \ldots, D$ as follows.
   - Generate an index $\ell^{(i)^*}$ from the set $\{1, \ldots, L\}$ with probability proportional to $\alpha_{\ell}^{\ddagger}$.
   - Generate $Z_{\ell^{(i)^*},t+1}^{(i)^*} \sim Q\left(z \,\middle|\, u_{t+1}^{*}\right)$, where $Q(z \mid v)$ is defined by (6.25).
   - Generate $Z_{\ell,t+1}^{(i)^*} \sim \Phi_1(x)$, for $\ell \neq \ell^{(i)^*}$.
   - Combine $Z_{\ell,t+1}^{(i)}$, $\ell = 1, \ldots, L$ with $Z_{\ell,t}^{(i)}, Z_{\ell,t-1}^{(i)}, \ldots, Z_{\ell,t-K+1}^{(i)}$ to obtain $x_{t+1,d}^{(i)}$, $d = 1, \ldots, D$ from the state equation (6.45).
3. For $i = 1, \ldots, N$, evaluate the importance weights $w_{t+1}^{(i)}$ according to (6.51).

CHAPTER 7

Estimation of Max-Stable Processes Using MCMC Methods

Markov chain Monte Carlo (MCMC) methods are now widely used in statistics, and the literature has grown enormously in the last years. In this Chapter we propose MCMC algorithms for estimating the posterior distributions of the unknown parameters and states in the state-space representation of M4 processes.

This Chapter is organized as follows. In §7.1 we review the background for MCMC methods. In §7.2 and §7.3 we provide full details for estimation of Moving Maxima (MM) processes. We outline the methods for estimation of M3 and M4 processes in §7.4.

Throughout this Chapter, we use the notation $x_{1:t} = (x_1, \ldots, x_t)'$ for $t > 1$, and $x_{\setminus j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_t)'$ to denote the vector $x_{1:t}$ without the jth component.

7.1 Markov Chain Monte Carlo Methods

In this section we present the basic background of MCMC methods, following closely Robert and Casella (2004). For a comprehensive treatment of the subject we refer to Gilks, Richardson and Spiegelhalter (1996), Robert and Casella (2004), and Gelman, Carlin, Stern and Rubin (1995).

A *Markov chain Monte Carlo (MCMC) method* to simulate from a distribution f is any method producing an ergodic Markov chain $\left\{x^{(i)},\ i = 1, 2, \ldots\right\}$ whose stationary distribution is f. Here we consider two well known MCMC methods: the *Metropolis-Hastings algorithm* (Metropolis *et al.*, 1953; Hastings, 1970), and the *Gibbs sampler* (Geman and Geman, 1984; Gelfand and Smith, 1990).

7.1.1 The Metropolis-Hastings Algorithm

The goal of the Metropolis-Hastings algorithm is to simulate from a *target distribution* with density f, based on draws from a *proposal distribution* with density $q(y \mid x)$. The algorithm is described by the following transition from $x^{(i)}$ to $x^{(i+1)}$:

Algorithm 7.1.1 (Metropolis-Hastings Algorithm).
Given $x^{(i)}$,

1. Generate $y \sim q\left(y \mid x^{(i)}\right)$.
2. Take
$$x^{(i+1)} = \begin{cases} y & \text{with probability } \rho(x^{(i)}, y), \\ x^{(i)} & \text{with probability } 1 - \rho(x^{(i)}, y), \end{cases}$$
where
$$\rho(x, y) = \min\left\{\frac{f(y)}{f(x)}\frac{q(x \mid y)}{q(y \mid x)},\ 1\right\}. \tag{7.1}$$

7.1.2 The Gibbs Sampler

The Gibbs sampler is an alternative form of the Metropolis-Hastings algorithm. The Gibbs sampler is particularly important when the goal is to simulate from a multivariate distribution. Following Robert and Casella (2004), suppose that the random vector $\boldsymbol{X}$ can be written as $\boldsymbol{X} = (X_1, \ldots, X_p)$ for some $p > 1$, where the X_j's are either univariate or multivariate. Suppose that we can simulate from the *full conditional densities*:

$$f_j(x_j \mid x_{\setminus j}), \qquad j = 1, \ldots, p.$$

The Gibbs sampling algorithm (or Gibbs sampler) is given in Algorithm 7.1.2.

7.1.3 Metropolis-within-Gibbs Algorithm

Sometimes it is not possible to simulate from one or more of the full conditional densities in the Gibbs sampler. In this case it is possible to use a Metropolis-Hastings step to substitute a draw from the full conditional density. This algorithm is called a *hybrid MCMC* or a *Metropolis-within-Gibbs* algorithm, and it is shown in Algorithm 7.1.3, which is adapted from Robert and Casella (2004, p. 393).

Algorithm 7.1.2 (The Gibbs Sampler).
Given $\boldsymbol{x}^{(i)} = \left(x_1^{(i)}, \ldots, x_p^{(i)}\right)$, generate

1. $x_1^{(i+1)} \sim f_1\left(x_1 \,\middle|\, x_2^{(i)}, \ldots, x_p^{(i)}\right)$;
2. $x_2^{(i+1)} \sim f_2\left(x_2 \,\middle|\, x_1^{(i+1)}, x_3^{(i)} \ldots, x_p^{(i)}\right)$;

   $\vdots$

p. $x_p^{(i+1)} \sim f_p\left(x_p \,\middle|\, x_1^{(i+1)}, \ldots, x_{p-1}^{(i+1)}\right)$.

Algorithm 7.1.3 (Metropolis-within-Gibbs Algorithm).
For $j = 1, \ldots, p$, given $\left(x_{1:j-1}^{(i+1)}, x_{j:p}^{(i)}\right) = \left(x_1^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_j^{(i)}, \ldots, x_p^{(i)}\right)$:

1. Generate $x_j^* \sim q_j\left(x_j \,\middle|\, x_{1:j-1}^{(i+1)}, x_{j:p}^{(i)}\right)$.
2. Take
$$x_j^{(i+1)} = \begin{cases} x_j^* & \text{with probability } \rho(x_j^{(i)}, x_j^*), \\ x_j^{(i)} & \text{with probability } 1 - \rho(x_j^{(i)}, x_j^*), \end{cases}$$
where
$$\rho(x_j, y_j) = \min\left\{ \frac{f_j\left(y_j \,\middle|\, x_{1:j-1}^{(i+1)}, x_{j+1:p}^{(i)}\right)}{f_j\left(x_j \,\middle|\, x_{1:j-1}^{(i+1)}, x_{j+1:p}^{(i)}\right)} \frac{q_j\left(x_j \,\middle|\, y_j, x_{1:j-1}^{(i+1)}, x_{j+1:p}^{(i)}\right)}{q_j\left(y_j \,\middle|\, x_j, x_{1:j-1}^{(i+1)}, x_{j+1:p}^{(i)}\right)},\ 1 \right\}. \tag{7.2}$$

Remark. As pointed out in Robert and Casella (2004), it is important to note that the Metropolis-Hastings step is used to generate a *single* x_j^* to substitute a draw from the full conditional distribution f_j, as opposed to generate multiple draws to approximate f_j.

7.2 Estimation of MM Processes

Following the approach of Chapter 6, the estimation method is based on a state-space representation of MM processes, where the state equation is an unobserved MM process, and the observed process is a nonlinear transformation of the state with a small additive Gaussian noise, to make the process nondegenerate.

The state-space representation of the MM process is

$$X_t = \max_{|k|\leq K} \alpha_k Z_{t-k}, \qquad t \in \mathbb{N}, \tag{7.3}$$

$$Y_t = \mu + \psi \frac{X_t^{\xi} - 1}{\xi} + \epsilon_t, \qquad t \in \mathbb{N}, \tag{7.4}$$

where $\{Z_t\}$ is a sequence of iid unit Fréchet random variables, and $\{\epsilon_t\}$ is a sequence of iid normally distributed random variables with mean 0 and variance $\sigma^2 > 0$. The vector

$$\theta = (\alpha,\ \mu,\ \log\psi,\ \xi,\ \log\sigma)' \tag{7.5}$$

denotes all the parameters in this model, where $\alpha = (\alpha_{-K}, \ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots, \alpha_K)'$ satisfies $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, $\xi \in \mathbb{R}$, $\psi > 0$, and $\mu \in \mathbb{R}$.

The likelihood contribution of the observation given the state variable is

$$f\left(y_t \mid x_t, \theta\right) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{y_t - \tilde{x}_t}{\sigma}\right), & y_t > u, \\ \Phi\left(\frac{u - \tilde{x}_t}{\sigma}\right), & y_t \leq u, \end{cases} \tag{7.6}$$

where $\tilde{x}_t$ is given by

$$\tilde{x}_t = \mu + \psi \frac{x_t^{\xi} - 1}{\xi}, \tag{7.7}$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution function, respectively. It should be noted that the symbols $\Phi(\cdot)$ and $\Phi_1(\cdot)$ are not related at all.

We assume that given the state and the parameters, the observations are independent. Therefore the likelihood of the model is

$$L\left(\theta \mid y_{1:T}\right) = f\left(y_{1:T} \mid x_{1:T}, \theta\right) = \prod_{t=1}^{T} f\left(y_t \mid x_t, \theta\right). \tag{7.8}$$

GPD-Based Parameterization

All of the methods developed in this Chapter can be written in terms of the GPD model for exceedances over thresholds. The only difference is the marginal transformation in the observation equation.

The GPD model can be written in terms of $\lambda = P\{Y > u\}$ and

$$P\{Y > y \mid Y > u\} = \left(1 + \xi \frac{y-u}{\beta}\right)_+^{-1/\xi},$$

so that

$$P\{Y \leq y\} = 1 - \lambda \left(1 + \xi \frac{y-u}{\beta}\right)_+^{-1/\xi}. \tag{7.9}$$

If X has unit Fréchet distribution, then for $X \geq -1/\log(1-\lambda)$ the transformation

$$Y = u + \frac{\beta}{\xi}\left[\left(\frac{1 - e^{-1/X}}{\lambda}\right)_+^{-\xi} - 1\right] \tag{7.10}$$

induces a random variable Y with distribution (7.9). Hence, a state-space representation of the MM process based on the GPD parameterization is obtained by replacing the marginal transformation (6.1) with (7.10).

7.2.1 Target Distribution

In a Bayesian framework it is of interest to obtain the joint posterior distribution of the states $Z_{1:T}$ and the parameter θ. The target distribution is

$$f(z_{1:T}, \theta \mid y_{1:T}) \propto f(y_{1:T} \mid x_{1:T}, \theta)\, f(z_{1:T} \mid \theta)\, f(\theta) \tag{7.11}$$

$$\propto f(y_{1:T} \mid x_{1:T}, \theta)\, f(z_{1:T})\, f(\theta). \tag{7.12}$$

We sample from the target distribution using a Gibbs sampler that iterates between the states and parameters. The parameter is divided into three components: the MM parameter α, the GEV parameters $(\mu, \log\psi, \xi)$, and the noise parameter $\log\sigma$. Given the states, each of these components is updated using the hybrid MCMC (Algorithm 7.1.3). On the other hand, given the parameters, the states are updated using an algorithm that combines single-move and block-move updating of the states, as proposed in Shephard and Pitt (1997).

The general structure of our Gibbs sampler is to iterate between the following steps:

1. Generate $Z_{1:T}^{(i+1)}$ from $f\left(z_{1:T} \mid \theta^{(i)}, y_{1:T}\right)$.

2. Generate $\alpha^{(i+1)}$ from $f\left(\alpha \,\middle|\, y_{1:T}, z_{1:T}^{(i+1)}, \mu^{(i)}, \psi^{(i)}, \xi^{(i)}, \sigma^{(i)}\right)$.

3. Generate $\mu^{(i+1)}, \psi^{(i+1)}, \xi^{(i+1)}$ from $f\left(\mu, \psi, \xi \,\middle|\, y_{1:T}, z_{1:T}^{(i+1)}, \alpha^{(i+1)}, \sigma^{(i)}\right)$.

4. Generate $\sigma^{(i+1)}$ from $f\left(\sigma \,\middle|\, y_{1:T}, z_{1:T}^{(i+1)}, \alpha^{(i+1)}, \mu^{(i+1)}, \psi^{(i+1)}, \xi^{(i+1)}\right)$.

In the following sections we discuss the updating steps in detail.

7.2.2 Updating MM Parameters

In this step the parameter of interest is $\alpha = (\alpha_{-K}, \ldots, \alpha_0, \ldots, \alpha_K)$. We assume a uniform (Dirichlet) prior on α with density

$$f(\alpha) = (2K)! \propto \text{constant}, \qquad \alpha_k \geq 0,\ \sum_k \alpha_k = 1. \tag{7.13}$$

See §B.3 for details about the Dirichlet distribution. The full conditional distribution in this step is

$$\begin{aligned} f(\alpha \mid y_{1:T}, z_{1:T}, \mu, \psi, \xi, \sigma) &\propto f(y_{1:T} \mid x_{1:T}, \theta)\, f(\alpha) \\ &\propto f(y_{1:T} \mid x_{1:T}, \theta). \end{aligned} \tag{7.14}$$

In other words, from $z_{1:T}$ we calculate $x_{1:T}$, which depends on α, and then evaluate the likelihood. Since it is not possible to sample from this distribution directly, we update α using a Metropolis-Hasting step where the proposal distribution has density of the form

$$q\big(\alpha \,\big|\, \alpha^{(i)}\big), \qquad \alpha_k \geq 0,\ \sum_k \alpha_k = 1. \tag{7.15}$$

Suppose at iteration i we have current estimates $Z_{1:T}^{(i+1)}$, $\theta^{(i)}$. Our approach is to generate proposals for each component α_k from a uniform distribution centered at $\alpha_k^{(i)}$, i.e.,

$$\alpha_k^* \sim \mathcal{U}\left(l_k^{(i)}, u_k^{(i)}\right), \qquad -K \leq k < K, \tag{7.16}$$

where

$$l_k^{(i)} = \max\left(0,\, \alpha_k^{(i)} - \varepsilon_\alpha\right), \tag{7.17}$$

$$u_k^{(i)} = \min\left(1,\, \alpha_k^{(i)} + \varepsilon_\alpha\right), \tag{7.18}$$

and ε_α is a tuning parameter. We repeat this procedure until $\sum_{k\neq K} \alpha_k^* \leq 1$, and then set $\alpha_K^* = 1 - \sum_{k\neq K} \alpha_k^*$. In general, ε_α should be small in order to accept proposals more often. However, choosing this value is a trial-and-error process.

The density of the kth component of α is

$$q\big(\alpha_k \,\big|\, \alpha_k^{(i)}\big) = \frac{1}{u_k^{(i)} - l_k^{(i)}}, \tag{7.19}$$

so the density of the proposal distribution is

$$q\big(\alpha^* \,\big|\, \alpha^{(i)}\big) = \prod_{k=-K}^{K-1} \frac{1}{u_k^{(i)} - l_k^{(i)}}. \tag{7.20}$$

Then we have

$$\frac{q\big(\alpha^{(i)} \,\big|\, \alpha^*\big)}{q\big(\alpha^* \,\big|\, \alpha^{(i)}\big)} = \prod_{k=-K}^{K-1} \frac{u_k^{(i)} - l_k^{(i)}}{u_k^* - l_k^*}. \tag{7.21}$$

Therefore, the probability of acceptance for the Metropolis-Hastings step is

$$\begin{aligned}
\rho\big(\alpha^{(i)}, \alpha^*\big) &= \min\left\{ \frac{f\big(y_{1:T} \,\big|\, x_{1:T}^*, \theta^*\big)}{f\big(y_{1:T} \,\big|\, x_{1:T}^{(i)}, \theta^{(i)}\big)} \frac{q\big(\alpha^{(i)} \,\big|\, \alpha^*\big)}{q\big(\alpha^* \,\big|\, \alpha^{(i)}\big)},\ 1 \right\} \\
&= \min\left\{ \prod_{t=1}^{T} \frac{f\big(y_t \,\big|\, x_t^*, \theta^*\big)}{f\big(y_t \,\big|\, x_t^{(i)}, \theta^{(i)}\big)} \prod_{k=-K}^{K-1} \frac{u_k^{(i)} - l_k^{(i)}}{u_k^* - l_k^*},\ 1 \right\}.
\end{aligned} \tag{7.22}$$

7.2.3 Updating GEV/GPD Parameters

We assume uniform distributions over a large interval as priors for each of the GEV parameters $(\mu, \log\psi, \xi)$. The full conditional distribution in this step is

$$\begin{aligned}
f\left(\mu, \log\psi, \xi \mid y_{1:T}, z_{1:T}, \alpha, \sigma\right) &\propto f\left(y_{1:T} \mid x_{1:T}, \theta\right) f\left(\mu, \psi, \xi\right) \\
&\propto f\left(y_{1:T} \mid x_{1:T}, \theta\right).
\end{aligned} \tag{7.23}$$

We sample from this distribution using a Metropolis step. For each of the GEV parameters we generate a proposal from a uniform distribution centered at the current value of the parameter. For the ξ parameter we reject any proposal that lies outside the interval $[-1, 1]$. For instance, given $\mu^{(i)}$ generate a proposal $\mu^* \sim \mathcal{U}\left(\mu^{(i)} - \varepsilon_\mu,\ \mu^{(i)} + \varepsilon_\mu\right)$, where ε_μ is a tuning parameter. Since the proposal density is symmetric, and the prior is uniform, the probability of acceptance for a Metropolis step is

$$\min\left\{\prod_{t=1}^{T} \frac{f\left(y_t \,\middle|\, x_t^*, \theta^*\right)}{f\left(y_t \,\middle|\, x_t^{(i)}, \theta^{(i)}\right)},\ 1\right\}. \tag{7.24}$$

The results from a large number of simulations suggest that the algorithm does not converge using the GEV parameterization. Hence, our approach is to work directly with the model in terms of the GPD parameters (λ, β, ξ) as defined by (7.10), and update the parameters using small uniform perturbations, as explained above.

An alternative approach (only discussed here for reference) was suggested in Coles (2003) for modeling daily rainfall data. Coles (2003) suggested transforming the GEV parameters as follows.

$$\lambda = 1 - \exp\left\{-\frac{1}{366}\left(1 + \xi\frac{u-\mu}{\psi}\right)^{-1/\xi}\right\}, \tag{7.25}$$

$$\beta = \log\left\{\psi + \xi(u - \mu)\right\}. \tag{7.26}$$

Taking a first order expansion of (7.25) and (7.26) gives back the usual parameterization of the GPD model, see (2.12). The number 366 is chosen because there are usually 366 daily observations per year for rainfall data. For financial data we would take 252.

7.2.4 Updating Noise Parameter

We assume a uniform distribution over a very large interval as prior for the noise parameter $\log \sigma$. The updating is performed in the same way as for the scale parameter in §7.2.3.

Alternative Updating

An alternative updating scheme for the σ parameter requires two steps: 1. Imputing the censored observations. 2. Updating using a conjugate prior. In this case the natural prior for σ^2 is an inverse gamma distribution. The inverse gamma distribution is denoted $\mathcal{IG}(a, b)$, and the corresponding density is

$$f(\sigma) = \frac{1}{\Gamma(a)b^a} \frac{1}{\sigma^{a+1}} \exp\left\{-\frac{1}{b\sigma}\right\}, \qquad \sigma \geq 0. \tag{7.27}$$

For a diffuse prior, we specify shape parameter $a = 0.001$ and scale parameter $b = 1/a$.

Given current values of states and parameters, we can generate the censored values from a normal distribution with mean $\widetilde{x}^{(i+1)}$ and standard deviation $\sigma^{(i)}$. Then given the observed and imputed data σ can be updated by a standard analytical method. Assuming that the prior is given by (7.27), the posterior of σ^2 is

$$\mathcal{IG}\left(a + \frac{n-1}{2}, \left[\frac{n-1}{2}s^2 + \frac{1}{b}\right]\right),$$

where s^2 is the sample standard deviation of all observed and imputed data. Then the updated value of σ for the Gibbs sampler is just the posterior mean.

7.3 Updating States of MM Processes

In this section we describe the methods for updating the states of MM processes given the parameters.

There are two main difficulties for designing an efficient sampler for updating the states. First, the state-space representation of MM processes is only valid for observations above a high-threshold, so even if the original sample size is large, we have limited information for extreme events. Second, the state is very high-dimensional and the components are dependent. Hence the sampler should take into account the time-series structure of the states.

In the stochastic volatility literature it is common to see that the volatility process is updated with the forward filtering backward sampling algorithm originally proposed by

Carter and Kohn (1994) and described in West and Harrison (1997). However, alternative approaches are needed when the state is inherently non-linear. Shephard and Pitt (1997) and Pitt and Shephard (1999b) propose methods that are possible to extend to non-linear state-space, and we follow closely their approach.

There is a trade-off between updating all the components of the state vector at the same time, or updating one component at a time. The former approach generates proposals that are accepted with very small probability, so there is small autocorrelation, but the chain moves very slowly and does not converge. On the other hand, updating one component at a time moves the chain, but the chain has high autocorrelation, hence convergence is also questionable. An intermediate approach is to consider blocks of observations, and update all of the corresponding states at the same time. Following Shephard and Pitt (1997), a *single-move sampler* updates one component of the state vector at a time. A *multi-move (or block) sampler* updates all of the components in a block of contiguous states at the same time.

The state vector for the MM process defined by equation (7.3) is $Z_{(-K+1):(T+K)} = (Z_{-K+1}, \ldots, Z_0, Z_1, \ldots, Z_{T+K})$. To simplify the exposition, and without loss of generality, in what follows we refer to the state vector as $Z_{1:T} = (Z_1, \ldots, Z_T)$.

7.3.1 Target Distribution

The natural prior for the state vector $Z_{1:T}$ is a product of unit Fréchet densities, i.e.,

$$f\left(z_{1:T}\right) = \prod_{t=1}^{T} \varphi_1(z_t) = \prod_{t=1}^{T} z_t^{-2} \exp\left\{-1/z_t\right\}. \tag{7.28}$$

Then the full conditional distribution of the states given the data and parameters is

$$\begin{aligned} f\left(z_{1:T} \mid y_{1:T}, \theta\right) &\propto f\left(y_{1:T} \mid x_{1:T}, \theta\right) f\left(z_{1:T}\right) \\ &= \prod_{t=1}^{T} f\left(y_t \mid x_t, \theta\right) \varphi_1(z_t). \end{aligned} \tag{7.29}$$

7.3.2 Proposal Distributions

The design of a proposal distribution for the states is a critical part of the sampling algorithm, especially for observations over a high threshold.

From Definition 3.6.1 recall that for a given integer K and a positive threshold u, we say that x_t is a *local maximum* if $x_t \geq \max\{u, x_{t-K:t+K}\}$.

In what follows, and in subsequent sections, we frequently refer to the ratio of the prior density of the states (φ_1, the unit Fréchet density) and the proposal density (q) that is used to generate candidate moves. It is convenient to define

$$\varpi(x, y) = \frac{\varphi_1(y)}{q(y \mid x)}. \tag{7.30}$$

Importance Sampling

Given $Z_t^{(i)}$ and the current observation y_t, our approach is to generate from different proposal densities according to the following:

- If y_t is not a local maximum[1] then generate a proposal $Z_t^* \sim \Phi_1(x)$. Then we have $\varpi\big(z_t^{(i)}, z_t^*\big) = 1$.

- If y_t is a local maximum then generate a proposal from an importance density. This is essentially the same approach we use in §6.2.4: If y_t is a local maximum, then we take the current $Z_t^{(i)}$ as the median of the distribution function of a unit Fréchet random variable given that it has exceeded a high threshold $u_t^{(i)}$, where $u_t^{(i)}$ is calculated according to (6.32). Then we generate $Z_t^* \sim Q\big(z \,\big|\, u_t^{(i)}\big)$ as defined in (6.25), and by (6.26) we have

$$\begin{aligned} \varpi\big(z_t^{(i)}, z_t^*\big) &= \frac{\varphi_1(z_t^*)}{q\big(z_t^* \,\big|\, u_t^{(i)}\big)} \\ &= h\big(z_t^*, u_t^{(i)}, \gamma\big) \end{aligned} \tag{7.31}$$

[1] These are mostly observations below the threshold, which are treated as censored. However there are also observations above the threshold that are not local maxima.

Following this approach, we can write for any t

$$\varpi\big(z_t^{(i)}, z_t^*\big) = \begin{cases} h\big(z_t^*, u_t^{(i)}, \gamma\big), & y_t \geq \max\left\{u, y_{t-K:t+K}\right\}, \\ 1, & \text{otherwise.} \end{cases} \tag{7.32}$$

Then the Metropolis-Hastings step for updating a single component Z_t is as follows. Given $Z_t^{(i)}$ and y_t, generate Z_t^* as described above. Then take

$$Z_t^{(i+1)} = \begin{cases} Z_t^* & \text{with probability } \rho\big(Z_t^{(i)}, Z_t^*\big), \\ Z_t^{(i)} & \text{with probability } 1 - \rho\big(Z_t^{(i)}, Z_t^*\big), \end{cases}$$

where

$$\rho\big(z_t^{(i)}, z_t^*\big) = \min\left\{ \prod_{j=t-K}^{t+K} \frac{f\big(y_j \,\big|\, x_j^*, \theta^{(i)}\big)}{f\big(y_j \,\big|\, x_j^{(i)}, \theta^{(i)}\big)} \frac{\varpi\big(z_t^{(i)}, z_t^*\big)}{\varpi\big(z_t^*, z_t^{(i)}\big)},\ 1 \right\}. \tag{7.33}$$

Remark. It is important to note that updating the component Z_t has an effect on the likelihood contribution of all the observations $y_{t-K:t+K}$. Furthermore, $f\left(y_t \,|\, x_t, \theta\right)$ is a function of $Z_{t-K:t+K} = (Z_{t-K}, \ldots, Z_t, \ldots, Z_{t+K})$. Hence, the evaluation of (7.33) requires implicitly that $Z_{t-K:t+K}^{(i)}$ are also given.

Uniform Proposal

A second approach for updating states at local maxima is to generate proposals from a uniform distribution centered at the current value of the state.

Given $Z_t^{(i)}$ and the current observation y_t, generate proposals according to the following:

- If y_t is not a local maximum then generate a proposal $Z_t^* \sim \Phi_1(x)$. Then we have $\varpi\big(z_t^{(i)}, z_t^*\big) = 1$.

- If y_t is a local maximum then generate $Z_t^* \sim \mathcal{U}\left(Z_t^{(i)} - \varepsilon_z,\ Z_t^{(i)} + \varepsilon_z\right)$, where ε_z is a tuning parameter. Since the proposal density is

$$q\big(z^* \,\big|\, z^{(i)}\big) = (2\varepsilon_z)^{-1}, \qquad \left|z^* - z^{(i)}\right| \leq \varepsilon_z, \tag{7.34}$$

in this case we have

$$\begin{aligned}\varpi\big(z_t^{(i)}, z_t^*\big) &= \frac{\varphi_1(z_t^*)}{q\big(z_t^* \,\big|\, z_t^{(i)}\big)} \\ &= 2\varepsilon_z \left(z_t^*\right)^{-2} \exp\left\{-1/z_t^*\right\}. \end{aligned} \tag{7.35}$$

Following this approach, we can write for any t

$$\varpi\big(z_t^{(i)}, z_t^*\big) = \begin{cases} 2\varepsilon_z \left(z_t^*\right)^{-2} \exp\left\{-1/z_t^*\right\}, & y_t \geq \max\left\{u, y_{t-K:t+K}\right\}, \\ 1, & \text{otherwise.} \end{cases} \tag{7.36}$$

The corresponding Metropolis-Hastings step is the same as before.

One-Sided vs. Two-Sided Models

There are some minor differences when considering a one-sided model ($\alpha_k = 0$ for $k < 0$) or a two-sided model ($\alpha_k > 0$ for $k < 0$). Suppose a single state Z_t is updated at a time. Updating Z_t has an effect on the likelihood contribution of a block of observations that depend on Z_t. Table 7.1 shows the observations and states that are to be considered when updating the state Z_t, for both one-sided and two-sided models.

	$t = 1, \ldots, K$	$t = K+1, \ldots, T-K$	$t = T-K+1, \ldots, T$
one-sided	$y_{t:t+K}$ $z_{t-K:t+K}$	$y_{t:t+K}$ $z_{t-K:t+K}$	$y_{t:T}$ $z_{t-K:T}$
two-sided	$y_{1:t+K}$ $z_{-K+1:t+2K}$	$y_{t-K:t+K}$ $z_{t-2K:t+2K}$	$y_{t-K:T}$ $z_{t-2K:T+K}$

Table 7.1: *Observations and States for a Single-Move Update.*

7.3.3 Multi-Move Sampler

We have mentioned that there is a trade-off between updating all of the components of the state vector at the same time, or updating one component at a time. An alternative algorithm is to consider blocks of contiguous states and update all of the components in a block at the same time. Following Shephard and Pitt (1997) this approach is called a *multi-move (or block) sampler.*

Stochastic Knots

According to Shephard and Pitt (1997) a *stochastic knot* is a time period that separates one block from another. The knots are stochastic because they are generated randomly at each iteration of the Gibbs sampler.

The blocks are of the form $(Z_{b_i+1}, Z_{b_i+2}, \ldots, Z_{b_{i+1}})'$, where $\{b_i,\ i = 1, \ldots, B\}$ are indices that determine the knots of the blocks. Hence, the multi-move sampler requires a tuning parameter B, which is the number of stochastic knots to use. Shephard and Pitt (1997) suggest selecting stochastic knots at times $b = (b_1, \ldots, b_B)'$, where

$$b_i = \left\lfloor T \times \frac{i + U_i}{B + 2} \right\rfloor, \qquad i = 1, \ldots, B, \tag{7.37}$$

T is the sample size, $U_i \sim \mathcal{U}(0, 1)$, and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. Then at each iteration of the Gibbs sampler, we generate the knots by (7.37), and update $B + 1$ blocks. A block can be visualized as a row of the matrix

$$\begin{array}{cccc}
Z_1, & Z_2, & \ldots, & Z_{b_1} \\
Z_{b_1+1}, & Z_{b_1+2}, & \ldots, & Z_{b_2} \\
Z_{b_2+1}, & Z_{b_2+2}, & \ldots, & Z_{b_3} \\
\vdots & \vdots & \vdots & \vdots \\
Z_{b_{B-1}+1}, & Z_{b_{B-1}+2} & \ldots, & Z_{b_B} \\
Z_{b_B+1}, & Z_{b_B+2}, & \ldots, & Z_T
\end{array}$$

Metropolis-Hastings Step

To simplify notation write $I_m = \{b_m + 1, b_m + 2, \ldots, b_{(m+1)}\}$, so for each $b_m,\ m = 1, \ldots, B$ the states to be updated are

$$Z_{I_m} = (Z_{b_m+1}, Z_{b_m+2}, \ldots, Z_{b_{(m+1)}})'.$$

We also need to consider beginning and end conditions, which are the vectors $Z^{(i)}_{b_m+1-2K:b_m}$ and $Z^{(i)}_{b_{(m+1)}+1:b_{(m+1)}+2K}$, respectively.

The strategy for our multi-move sampler is as follows. Given $Z_t^{(i)}$, and beginning and end conditions, generate a proposal Z_t^* for each Z_t, $t \in I_m$ according to the methods described in §7.3.2. Then take

$$Z_{I_m}^{(i+1)} = \begin{cases} Z_{I_m}^* & \text{with probability } \rho\big(Z_{I_m}^{(i)}, Z_{I_m}^*\big), \\ Z_{I_m}^{(i)} & \text{with probability } 1 - \rho\big(Z_{I_m}^{(i)}, Z_{I_m}^*\big), \end{cases}$$

where

$$\rho\left(z_{I_m}^{(i)}, z_{I_m}^*\right) = \min\left\{ \prod_{j=b_m+1-K}^{b_{(m+1)}+K} \frac{f\big(y_j \,\big|\, x_j^*, \theta^{(i)}\big)}{f\big(y_j \,\big|\, x_j^{(i)}, \theta^{(i)}\big)} \frac{\varpi\big(z_j^{(i)}, z_j^*\big)}{\varpi\big(z_j^*, z_j^{(i)}\big)},\ 1 \right\}. \tag{7.38}$$

Combining Single-Move and Multi-Move Samplers

In practice we combine single-move updates with multi-move updates. This approach is suggested by Shephard and Pitt (1997) because sometimes the multi-move sampler can move slowly in the state-space, so in order to guarantee that the sampler moves, it is suggested to use the multi-move sampler, and perform a single-move update at every k, say 10, iteration of the Gibbs sampler.

7.4 Estimation of M3 and M4 Processes

In this section we outline extensions of the algorithms developed for estimation of MM processes to the case of M3 and M4 processes.

7.4.1 Updating M3 Parameters

The parameter for M3 processes is a matrix of L rows and $2K+1$ columns:

$$\alpha = \begin{bmatrix} \alpha_{1,-K} & \cdots & \alpha_{1,-1} & \alpha_{1,0} & \alpha_{1,1} & \cdots & \alpha_{1,K} \\ \alpha_{2,-K} & \cdots & \alpha_{2,-1} & \alpha_{2,0} & \alpha_{2,1} & \cdots & \alpha_{2,K} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{L,-K} & \cdots & \alpha_{L,-1} & \alpha_{L,0} & \alpha_{L,1} & \cdots & \alpha_{L,K} \end{bmatrix}.$$

Assume a uniform prior on α, with density

$$f(\alpha) = \Gamma(L(2K+1)) \propto \text{constant}, \qquad \alpha_{\ell k} \geq 0, \ \sum_{\ell} \sum_{k} \alpha_{\ell k} = 1. \tag{7.39}$$

Then the full conditional distribution in this step is

$$f\left(\alpha \mid y_{1:T}, z_{1:T}, \theta_{\setminus \alpha}\right) \propto f\left(y_{1:T} \mid z_{1:T}, \theta\right) f(\alpha) \tag{7.40}$$

$$= f\left(y_{1:T} \mid z_{1:T}, \theta\right) \tag{7.41}$$

We sample from the full conditional distribution of α by using a Metropolis-Hasting step where the proposal distribution is uniform centered at the current value of α. Suppose at iteration i we have current estimates $Z_{1:T}^{(i+1)}$, $\alpha^{(i)}$, $\theta_{\setminus \alpha}^{(i)}$. We would like sample proposals for α from a distribution with density function of the form

$$q\big(\alpha \,\big|\, \alpha^{(i)}\big) = \prod_{\ell=1}^{L} \prod_{k=-K}^{K} q\big(\alpha_{\ell k} \,\big|\, \alpha_{\ell k}^{(i)}\big) \qquad \alpha_{\ell k} \geq 0, \ \sum_{\ell} \sum_{k} \alpha_{\ell k} = 1. \tag{7.42}$$

A simple proposal generates each component $\alpha_{\ell k}$ based on $\alpha_{\ell k}^{(i)}$, but independently of the other α's. If the generated α's do not add up to 1 then we reject. This is the same method used for MM processes.

If all the entries of α are updated at once, there will be many rejections in the Metropolis-Hastings step because α is very high-dimensional. A sensible strategy is to update two entries of the matrix at a time and then take a Metropolis step. One could randomly select two entries or process the entries in order. Another sensible blocking strategy is to consider updating one row of the matrix at a time.

7.4.2 Updating States of M3 Processes

The updating of states is based on the importance density defined for local maxima in Algorithm 6.3.4. The idea is that if y_t is a local maximum, then it defines a candidate signature pattern, and we need to identify which of the patterns is driving the process for that particular time period.

The states of M3 processes can be represented as a matrix of L rows and $T + 2K + 1$

columns:

$$Z = \begin{bmatrix} Z_{1,-K+1} & \cdots & Z_{1,-1} & Z_{1,0} & Z_{1,1} & \cdots & Z_{1,T} & \cdots & Z_{1,T+K} \\ Z_{2,-K+1} & \cdots & Z_{2,-1} & Z_{2,0} & Z_{2,1} & \cdots & Z_{2,T} & \cdots & Z_{2,T+K} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ Z_{L,-K+1} & \cdots & Z_{L,-1} & Z_{L,0} & Z_{L,1} & \cdots & Z_{L,T} & \cdots & Z_{L,T+K} \end{bmatrix}.$$

The proposals $\{Z^*_{\ell t},\ \ell = 1, \ldots, L\}$ for updating the states of M3 processes at time t are generated as follows.

- If y_t is a local maximum then

  1. generate an index $\ell^* \in \{1, \ldots, L\}$, where each $\ell \in \{1, \ldots, L\}$ can be selected with probability $\alpha^+_\ell = \sum_k \alpha_{\ell k}$;
  2. given ℓ^*, generate $Z^*_{\ell^*,t}$ using the methods described in §7.3.2;
  3. given ℓ^*, generate $Z^*_{\ell t} \sim \Phi_1(x)$ for $\ell \neq \ell^*$.

- If y_t is not a local maximum then generate $Z^*_{\ell t} \sim \Phi_1(x)$ for $\ell = 1, \ldots, L$.

7.4.3 Updating States of M4 Processes

In analogy with the particle filtering method in §6.4, it is straightforward to extend the method to generate proposals for the states of M4 processes. Again, the minor differences from M3 processes are the definition of local maximum, which is given by Definition 3.6.5, the likelihood evaluations, and the relative frequency of the signature patterns.

7.5 Simulation Results

In this section we summarize simulation results for the MCMC algorithm for MM processes. The results presented are representative of our findings in many simulation experiments. The main goal in this section is to compare the estimated parameters and states with the actual values from the simulated process.

We simulate a sample path of $T = 3000$ observations from the model defined by (7.3) and (7.4). We set the parameter of the MM process to $(\alpha_{-1}, \alpha_0, \alpha_1) = (0.1, 0.7, 0.2)$. The

GEV marginal transformation has location, scale, and shape parameters $\mu = 0$, $\psi = 0.5$, and $\xi = 0.2$, respectively. The noise parameter is $\sigma = 0.2$. This model has a total of 6 parameters because of the restriction $\sum \alpha_k = 1$.

We assume that only the exceedances above the threshold $u = 1.7$ are observed. This threshold corresponds to the 0.925 empirical quantile of the simulated process and it is chosen arbitrarily. The results are not affected as long as we select a large enough threshold, say no smaller than the 0.9 empirical quantile of the simulated data.

Our strategy for initializing the chain is as follows. For the MM parameter use the fitted parameters from the clustering-based algorithm. For the GPD parameters we use the maximum likelihood estimates. For the noise parameter we use an arbitrary number smaller than the GPD scale parameter. For the states we use unit Fréchet draws for censored observations, and an approximation to the exceedance in unit Fréchet scale for the observed exceedances.

Note that the data is generated using the GEV as the marginal distribution, however the estimation is performed in terms of the GPD parameterization. This is possible because the two models are consistent above the threshold u, as discussed in Chapter 2. Many simulation experiments show divergence of the chain using the GEV parameterization, however the mixing properties of the chain seem much better with the GPD parameterization.

We let the chain run for a burn-in period of 20,000 iterations. After that, we let the chain run for another 100,000 iterations, but we only record observations every 10 iterations, so the final MCMC output is a sample of size 10,000.

The length of the uniform density that we use for the proposal distributions of the parameters is a tuning parameter that can be selected by trial-and-error. After a few experiments we select tuning parameters that produce a chain that seems to have reasonable mixing properties.

For updating the states corresponding to local maxima we use proposals from the exceedances over a threshold of a unit Fréchet, as described in §7.3.2. In order to prevent the chain to get stuck in any particular region of the state-space, we combine single-move and multi-move updating, i.e., at every iteration we use multi-move updating, and at every 10 iterations we perform a single-move updating. For the multi-move updating we select

$B = 30$ random knots, which is equivalent to updating on average 100 state variables at each time.

7.5.1 Monitoring Convergence

We analyze the MCMC output for the marginal posterior distributions of parameters and states. In order to analyze convergence to the posterior distributions we provide the usual convergence diagnostics:

- Acceptance rate, which is the number of times a proposal is accepted divided by the number of iterations. If the acceptance rate is too small then it indicates poor mixing of the chain. If the rate is too high then it indicates that the chain is only exploring a small region of the space.
- Time plot, which displays the mixing properties of the chain as a function of the number of iterations.
- Autocorrelation plot, which displays the autocorrelation function of the chain.
- Density plot, which displays the estimated posterior distribution based on a kernel density estimate of the chain.

Table 7.2 shows the acceptance rates for the parameters and states. Table 7.3 shows the actual values of the parameters, and estimated quantiles of the marginal posterior distributions of the parameters.

Figure 7.1 and Figure 7.2 show diagnostic plots for the parameters. The time plots show good mixing in all parameters. The autocorrelation plots include significant lags for very high lags, but they eventually decay. The density plots show the marginal posterior distribution of each parameter, with a vertical reference line indicating the actual value of the parameter. The actual value of the parameter is covered by the posterior distribution in all cases, except for the λ parameter.

Figure 7.3 shows the observed exceedances over the threshold, and the actual states that generated the process for the time period $521 \leq t \leq 550$. We choose this interval because it illustrates the estimation of a local maximum at y_{527}. Note that the observation

y_{528} exceeds the threshold, but it is not a local maximum, hence it is very likely that it was caused by the lagged effect of the shock z_{527}.

It is difficult to visualize convergence of the state vector because of its high-dimensionality. In order to get an idea of the convergence of states we select three states—corresponding to z_{526}, z_{527}, and z_{528}—and show diagnostic plots in Figure 7.4. Note that z_{527} is a very large shock and the posterior distribution captures this correctly.

Furthermore, we estimate the overall acceptance rate of states corresponding to local maxima to be 0.0304, while the acceptance rate for the rest of the observations is 0.1016.

Parameter	Acceptance Rate
α	0.0899
ξ	0.3026
β	0.1501
λ	0.0776
σ	0.3420

Table 7.2: *Acceptance Rates of Parameters from MCMC Output.*

Parameter	Actual	Quantiles				
		0.005	0.025	0.500	0.975	0.995
α_{-1}	0.1	0.0830	0.0876	0.1002	0.1122	0.1156
α_0	0.7	0.6726	0.6783	0.6988	0.7208	0.7263
α_1	0.2	0.1842	0.1881	0.2010	0.2139	0.2177
λ	0.075	0.0375	0.0419	0.0528	0.0622	0.0676
β	0.84	0.7365	0.7513	0.8263	0.9144	0.9493
ξ	0.2	0.1482	0.1622	0.2097	0.2601	0.2759
σ	0.2	0.1574	0.1646	0.1952	0.2360	0.2515

Table 7.3: *Quantiles of Posterior Distribution from MCMC Output. Actual values shown in first column.*

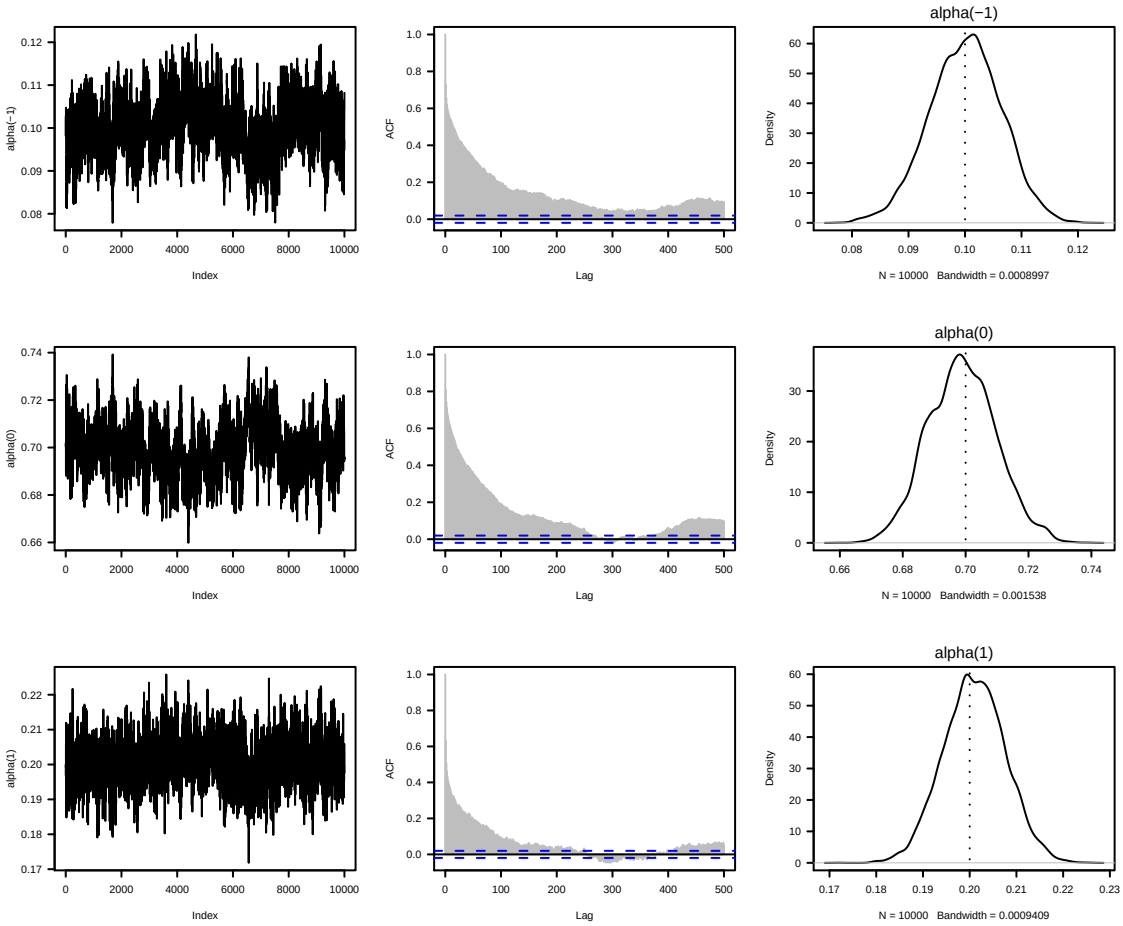

Figure 7.1: *MCMC Output for MM Process Parameters. From top to bottom:* α_{-1}, α_0, *and* α_1.

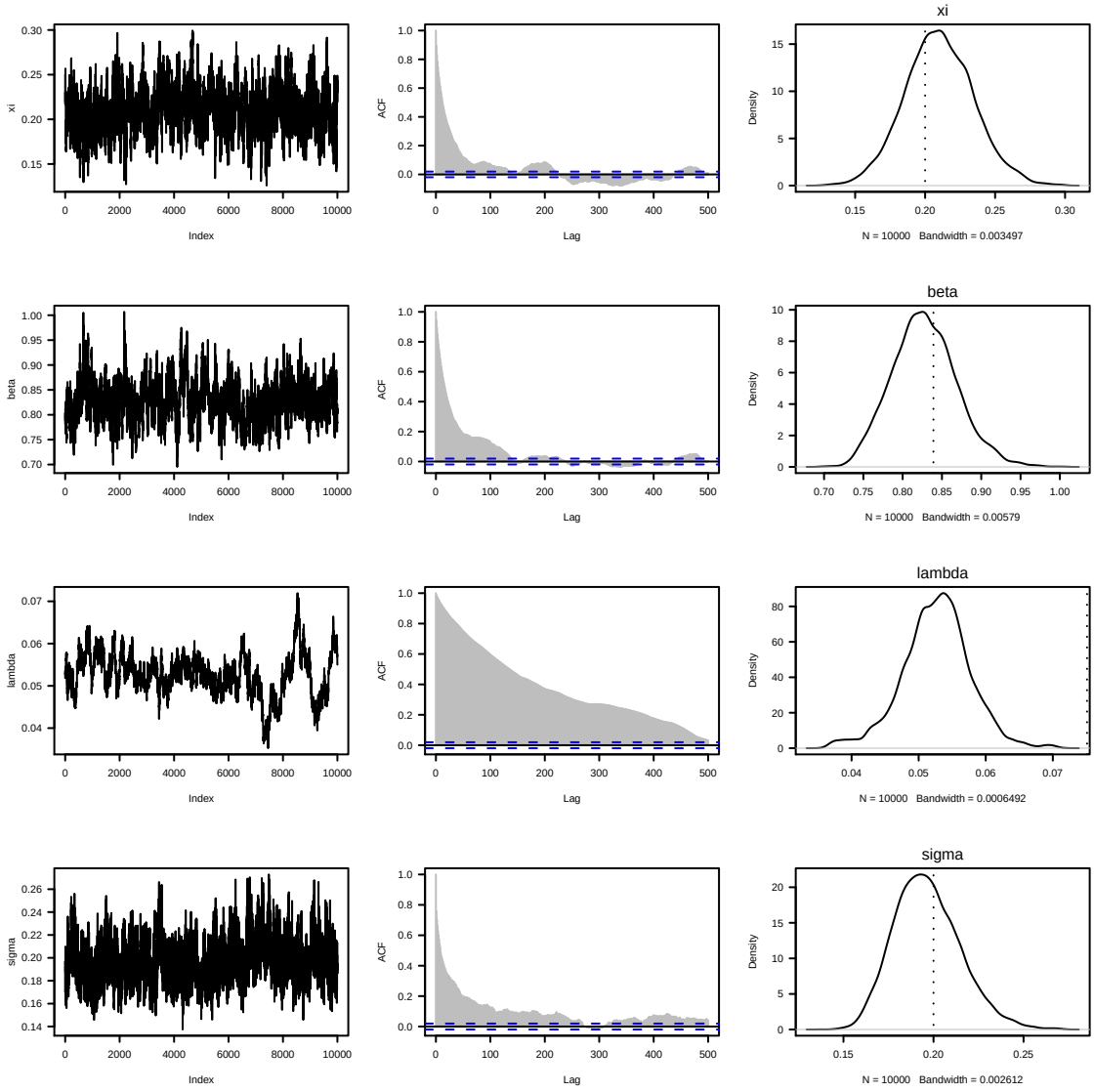

Figure 7.2: *MCMC Output for Extreme Value and Noise Parameters. From top to bottom:* ξ, β, λ, *and* σ.

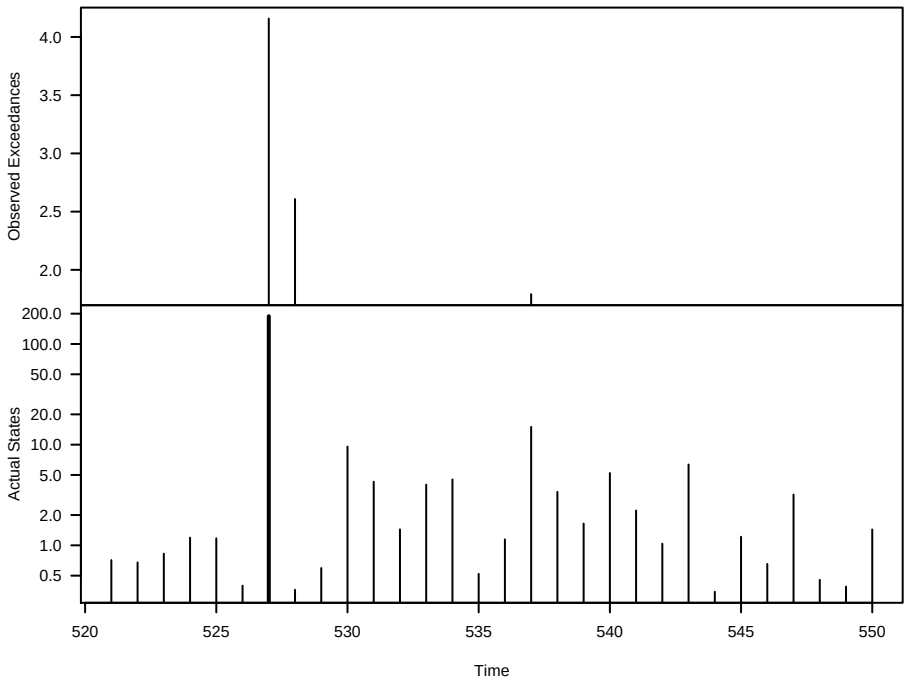

Figure 7.3: *Observed Exceedances and Actual States of Simulated MM Process. The largest shock is highlighted to remark that it induces two exceedances over the threshold.*

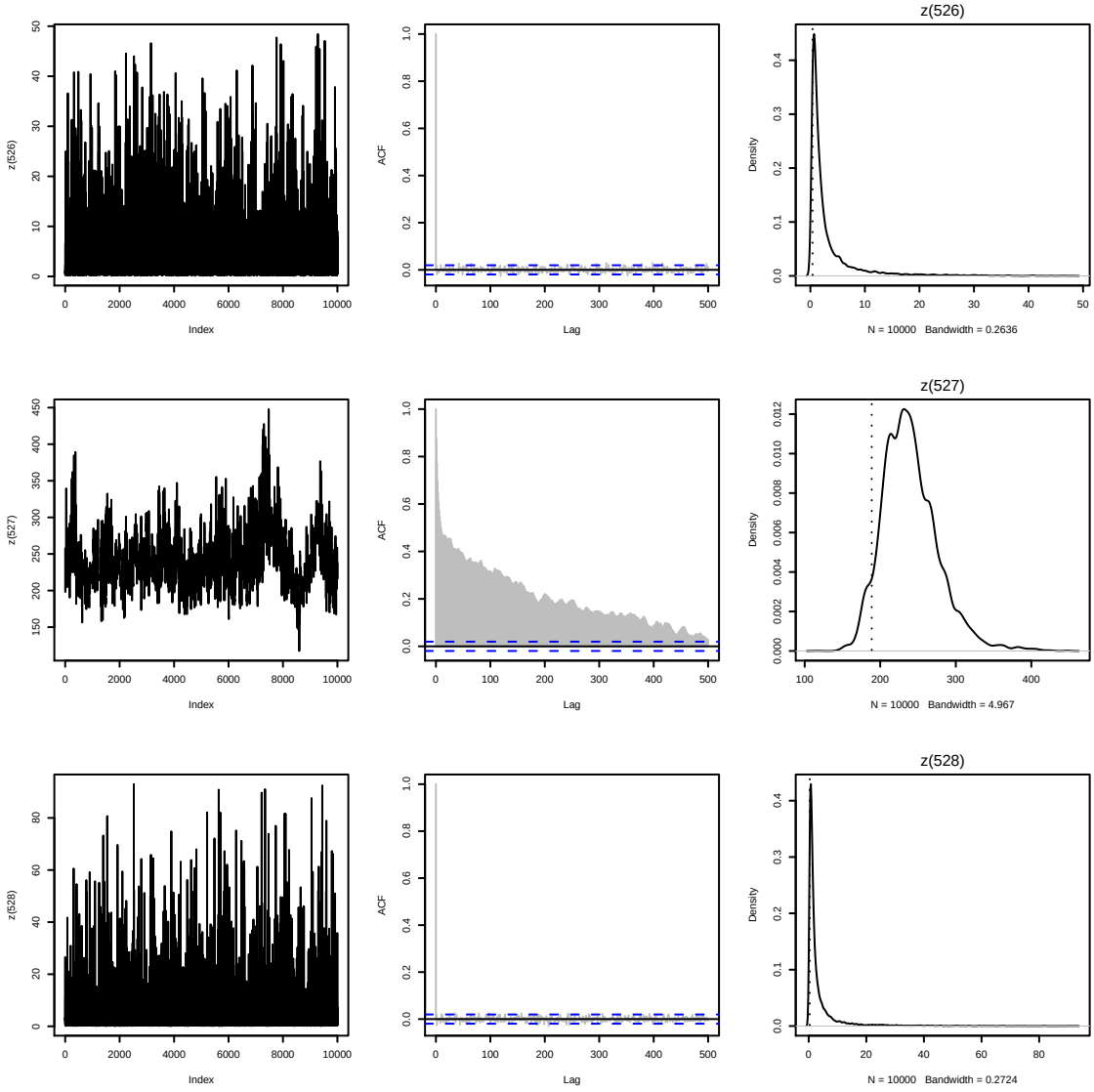

Figure 7.4: *MCMC Output for Selected States of Simulated MM process. From top to bottom:* z_{526}, z_{527}, *and* z_{528}.

CHAPTER 8

Financial Econometrics and Risk Assessment

In this Chapter we review the basic background for financial econometrics, with emphasis on volatility models. We discuss financial risks, and introduce common measures of market risk, namely Value at Risk (VaR) and Expected Shortfall (ES). We propose simulation-based prediction of M4 processes, and apply our methods to estimate market risk of a hypothetical portfolio of Nasdaq sector indices.

8.1 Financial Econometrics

Financial econometrics is concerned with the modeling of financial data using statistical (econometric) tools. Engle (2001) mentions that a central concern of financial econometrics is to discover the joint conditional density of asset prices observed at time t given the information set at time t which must include these prices. Usually the analysis is carried out in terms of log-returns because they have better statistical properties than prices.

The models in financial econometrics attempt to capture "stylized facts" of financial time series, such as time-varying volatility, non-normality and heavy-tailedness of the distribution of asset returns, and long-range dependence. We refer to Bollerslev, Engle and Nelson (1994) or Shephard (1996) for a more extensive discussion.

8.1.1 Returns and Portfolios

Following Chapter 2 of Campbell, Lo and MacKinlay (1997), suppose that P_t is an asset price at date t and assume this asset pays no dividends. The *log return* (or *continuously*

compounded return) is defined by

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right). \tag{8.1}$$

The log return over the most recent h periods is

$$\begin{aligned} r_t(h) &= \log\left(\frac{P_t}{P_{t-h}}\right) \\ &= \log\left(\frac{P_t}{P_{t-1}}\frac{P_{t-1}}{P_{t-2}}\cdots\frac{P_{t-h+1}}{P_{t-h}}\right) \\ &= r_t + r_{t-1} + \cdots + r_{t-h+1}. \end{aligned} \tag{8.2}$$

In words, the multi-period log return is the sum of the single-period log returns, so log returns are time-additive.

A portfolio is simply a collection of assets owned by an investor. The value of a portfolio, denoted $P_{t,p}$, composed of D assets at time t is

$$P_{t,p} = \sum_{d=1}^{D} w_d P_{t,d}, \tag{8.3}$$

where $P_{t,d}$ is the market price of asset d and w_d is the proportion invested in that asset. We assume $w_d > 0$ and $\sum w_d = 1$. Usually portfolios have a risk-free asset, such as a short-term Treasury bill. The log return of a portfolio is denoted $r_{t,p}$. It is important to note that $r_{t,p}$ is not the weighted sum of individual log returns, however, in practice the following approximation is commonly used

$$r_{t,p} \approx \sum_{d=1}^{D} w_d r_{t,d}. \tag{8.4}$$

8.2 Volatility Models

Time series models that allow for the modeling of time varying variances and covariances are called volatility models. Much of the research in volatility models has been motivated by empirical "stylized facts" of financial time series and the pricing of contingent assets.

The assumption of a continuous time model for option pricing is commonly used in financial econometrics. The main assumption is that the value of some underlying security, P, follows a geometric diffusion, e.g.

$$\mathrm{d}P = mP\mathrm{d}t + sP\mathrm{d}W, \tag{8.5}$$

where W is a standard *Brownian motion*.

An asset c whose value is a function of P is called *contingent* or *derivative*. The standard example is a European option. The classical theory for pricing European options is based on the *Black-Scholes* formulas. However, this pricing scheme relies on the assumption of constant variance, which is inconsistent with the data. This has motivated extensions to the Black-Scholes approach for option pricing, including the modeling of second-order processes with diffusions, and the incorporation of jumps, both in the mean and the volatility diffusions. A recent survey of the large literature on the subject is Garcia, Ghysels and Renault (2005).

Shephard (1996) adopts a classification that divides these models into "observation-driven" and "parameter-driven" models. Broadly speaking, this classification reduces to two groups: the class of *autoregressive conditional heteroskedastic* (ARCH) models and their generalizations, and *stochastic volatility* (SV) models. More recently, a new class of models, known as MIDAS models have been introduced.

8.2.1 ARCH Models

The ARCH class of models was introduced by Engle (1982). We follow closely the definition in Bollerslev *et al.* (1994). Let $\{\epsilon_t\}$ be a discrete time stochastic process in $\mathbb{R}$, and let $\mathcal{F}_{t-1}$ be the sigma-field generated by the past observations and any other information available at time $t-1$. Then $\{\epsilon_t\}$ is defined to follow an ARCH model if

$$\mathbb{E}\left(\epsilon_t \mid \mathcal{F}_{t-1}\right) = 0, \qquad t = 1, 2, \ldots, \tag{8.6}$$

but the conditional variance,

$$s_t^2 \equiv \mathbb{E}\left(\epsilon_t^2 \,\middle|\, \mathcal{F}_{t-1}\right), \qquad t = 1, 2, \ldots, \tag{8.7}$$

depends non-trivially on the sigma-field generated by the past observations. We consider ARCH models of the form:

$$r_t = m_t + s_t Z_t, \tag{8.8}$$

where

- $\{Z_t\}$ are iid with distribution function $F_Z(z)$, $\mathbb{E}(Z_t) = 0$, and $\mathbb{E}\left(Z_t^2\right) = 1$.
- m_t and s_t are $\mathcal{F}_{t-1}$-measurable; in other words, m_t and s_t depend non-trivially on the sigma-field generated by past observations.

The GARCH$(1,1)$ model introduced by Bollerslev (1986) is probably the most used in practice. This model is defined by (8.6) and

$$s_t^2 = \beta_0 + \beta_1 \epsilon_{t-1}^2 + \beta_2 s_{t-1}^2, \tag{8.9}$$

where $\beta_0 > 0$, $\beta_1 \geq 0$, and $\beta_2 \geq 0$. This process is strictly stationary if $\beta_1 + \beta_2 < 1$.

For a comprehensive survey on ARCH models refer to Bollerslev *et al.* (1994) and Shephard (1996).

8.2.2 SV Models

SV models are usually defined in a state-space framework. A simple example of an SV model given in Taylor (1986) is

$$Y_t = \epsilon_t \exp\{X_t/2\}, \tag{8.10}$$

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \eta_t, \tag{8.11}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$, $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, and ϵ_t, η_t are independent. Properties of this model are discussed in Taylor (1986) and Shephard (1996). The latter points out that the main problem with these models is the lack of analytic one-step-ahead forecasting densities.

However, these models have nice properties and it is easier to generalize them to the multivariate case. Estimation methods for these models include Generalized Method of Moments (Andersen and Sorensen, 1996), quasi-maximum likelihood (Harvey, Ruiz and Shephard, 1994), MCMC (Jacquier, Polson and Rossi, 1994), and particle filters (Pitt and Shephard, 1999a).

8.2.3 MIDAS Models

Ghysels, Santa-Clara and Valkanov (2004a,b) introduced *Mixed Data Sampling* (henceforth *MIDAS*) models for time-varying volatility. The MIDAS approach combines data sampled at different frequencies to predict volatility. For example, prediction of daily volatility is based on intra-daily returns.

The MIDAS models is a regression of the form

$$V_{t+h,t}^{(hm)} = \mu_h + \phi_h \sum_{k=0}^{k^{\max}} b_h(k,\theta) \tilde{X}_{t-k,t-k-1}^{(m)} + \varepsilon_{ht}, \tag{8.12}$$

where $\tilde{X}_{t-k,t-k-1}^{(m)}$ are the regressors, and the lag coefficients $b_h(k,\theta)$ are nonnegative and satisfy $\sum_k b_h(k,\theta) = 1$. A parsimonious specification is to assume $\theta = (\theta_1, \theta_2)$, and

$$b_h(k,\theta) \propto f\left(\frac{k}{k^{\max}}, \theta_1, \theta_2\right), \tag{8.13}$$

where $f(z,a,b)$ is the Beta function given by

$$f(z,a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} z^{a-1}(1-z)^{b-1}. \tag{8.14}$$

The choice of regressors is the object of current research, for example, past values of the absolute returns can be used (Ghysels *et al.*, 2004a). A MIDAS model of the form (8.8) based on daily absolute returns can be written

$$s_{t+1}^2 = \mu + \phi \sum_{k=0}^{k^{\max}} b(k,\theta) \left| r_{t-k} \right| + \varepsilon_{t+1}. \tag{8.15}$$

8.3 Financial Risks

Financial risk management has evolved rapidly during the last couple of decades. A number of factors combined have produced this boom, such as the growth of the financial markets, the introduction of derivatives, and recent financial disasters. Some of the largest financial institutions have suffered losses in the hundreds of millions of dollars in the financial markets, which are believed to be because of the misuse of derivative products and the lack of a good financial risk management system. This has increased the awareness of investors and regulators of the financial system to obtain accurate quantitative measures of financial risks. These measures are routinely used for internal risk management, for investment decisions and for regulation of financial institutions, among others.

One of the main concerns has been the measurement of *market risk*, which is defined as the possible loss due to fluctuations in the prices of the assets. The Bank of International Settlements (BIS, 1993) defines many other types of financial risks, such as *credit risk*, *operational risk*, and *systemic risk*. We refer to (BIS, 1993) for a detailed list of financial risks and their definitions. In this work we will only focus on market risk.

8.3.1 Coherent Measures of Risk

Artzner, Delbaen, Eber and Heath (1999) provide an axiomatic framework for risk management. Four of the axioms state the desirable properties that measures of risk should satisfy if they are to be used for setting capital requirements. Let X and Y be the future net worth of two market positions. Let r be a risk-free rate, $a \in \mathbb{R}$, and $\lambda > 0$. A measure of risk ρ is called *coherent* if it satisfies the following properties:

$$\rho(X+Y) \leq \rho(X) + \rho(Y) \qquad \text{(subadditivity)} \tag{8.16}$$

$$\rho(\lambda X) = \lambda \rho(X) \qquad \text{(homogeneity)} \tag{8.17}$$

$$\rho(X) \geq \rho(Y) \text{ if } X \leq Y \qquad \text{(monotonicity)} \tag{8.18}$$

$$\rho(X + ra) = \rho(X) - a \qquad \text{(risk-free condition)} \tag{8.19}$$

The subadditivity condition is a natural requirement that ensures that the measure of

risk behaves reasonably when adding positions. For instance, Artzner *et al.* (1999) argue that it would be unreasonable for an exchange to require more margin for the sum of two positions than the sum of the margins required for each individually. Both subadditivity and homogeneity imply that the function ρ is convex. The monotonicity condition implies that if $X \leq Y$, then an investor would prefer Y, so a coherent measure of risk has to assign less risk to Y. The risk-free condition states that if the amount a is added to a position, and invested at a risk-free rate r, then the risk is reduced in the amount of a.

Value at Risk (VaR) has become the most widely accepted tool to measure market risk, and it is now a standard in the industry. The so-called *Expected Shortfall* (ES) is an alternative measure with better theoretical properties.

8.3.2 Value at Risk and Expected Shortfall

Intuitively, the Value at Risk is the maximum loss that the value of an asset (or a portfolio of assets) can suffer with a given probability and during a specified time-horizon. According to the BIS regulations, the VaR is the capital that a bank needs to hold to cover possible losses on its trading portfolio over a 10 day period in 99% of the occasions. Formulated as a statistical problem, the VaR is defined as a quantile of the distribution of portfolio's return over the specified time horizon. Let $X = -r_{t+h}(h)$ be the negative of the cumulative log return over the time interval $(t, t+h)$. The conditional VaR is defined by

$$\mathrm{VaR}_{q,t+h}(h) = \inf\left\{x \in \mathbb{R} : F_{X|\mathcal{F}_t}(x) \geq q\right\}, \tag{8.20}$$

where q is a specified probability, usually 0.95 or 0.99.

The conditional Expected Shortfall is a tail conditional expectation, defined as the average of the negative returns given that they exceed the conditional VaR, i.e.,

$$\mathrm{ES}_{q,t+h}(h) = \mathbb{E}\left(X \mid X \geq \mathrm{VaR}_{q,t+h}(h),\, \mathcal{F}_t\right). \tag{8.21}$$

To simplify notation, when the time horizon is one day we denote these quantities as $\mathrm{VaR}_{q,t+1}$ and $\mathrm{ES}_{q,t+1}$, respectively.

Although VaR is very popular among practitioners, it has several disadvantages. For

instance, estimating the VaR at very high confidence levels is too volatile. On the other hand, the VaR only provides the probability of a "bad" event, but it does not tell how large that event is going to be. Furthermore, Artzner *et al.* (1999) have criticized the use of VaR because it is not subadditive, hence it is not a coherent measure of risk. Daníelsson *et al.* (2001) discuss other issues of using VaR for setting capital requirements.

On the other hand, the Expected Shortfall is a coherent measure and it is sensitive to very large events in the tail, which allows to distinguish between two distributions with the same quantile, but different tail.

8.3.3 Risk Assessment Based on ARCH-GPD Models

The approach of McNeil and Frey (2000) is to combine ARCH and GPD models for estimating VaR and ES. The relevance of the work by McNeil and Frey (2000) is that it provides explicit formulas for the estimation of VaR and ES, and it combines two important features of financial data in the model, namely time-varying volatility and heavy tails.

McNeil and Frey (2000) assume a model of the form (8.8), where m_t and s_t are an AR(1) process and a GARCH(1, 1) model, respectively. After fitting this model, the standardized residuals $z_t = (r_t - \hat{m}_t)/\hat{s}_t$ are assumed to be iid, but with unknown distribution function. The standardized residuals above a high threshold are then used in a second stage to fit a GPD model for the tails of the error distribution. Measures of risk are based on the one-step predictive distribution of the asset return, given by

$$\begin{aligned} P\{r_{t+1} \leq x \mid \mathcal{F}_t\} &= P\{s_{t+1}Z_{t+1} + m_{t+1} \leq x \mid \mathcal{F}_t\} \\ &= F_Z\left(\frac{x - m_{t+1}}{s_{t+1}}\right). \end{aligned} \tag{8.22}$$

Therefore the conditional VaR (quantile) is

$$\mathrm{VaR}_{q,t+1} = m_{t+1} + s_{t+1}z_q, \tag{8.23}$$

where z_q is the upper qth quantile of Z. The conditional ES is

$$\mathrm{ES}_{q,t+1} = m_{t+1} + s_{t+1}\mathbb{E}(Z \mid Z > z_q). \tag{8.24}$$

Based on the GPD assumption, the estimated conditional VaR is

$$\widehat{\mathrm{VaR}}_{q,t+1} = \hat{m}_{t+1} + \hat{s}_{t+1}\hat{z}_q, \tag{8.25}$$

and the estimated conditional ES is

$$\widehat{\mathrm{ES}}_{q,t+1} = \hat{m}_{t+1} + \hat{s}_{t+1}\hat{z}_q \left[\frac{1}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}\hat{z}_{(k+1)}}{(1-\hat{\xi})\hat{z}_q} \right], \tag{8.26}$$

where k is the index of the kth largest order statistic, and

$$\hat{z}_q = z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}} \left[\left(\frac{1-q}{k/n} \right)^{-\hat{\xi}} - 1 \right]. \tag{8.27}$$

Simulation approaches to obtain estimates of $\mathrm{VaR}_{q,t+h}(h)$ and $\mathrm{ES}_{q,t+h}(h)$, for $h > 1$ have been proposed independently by McNeil and Frey (2000) and Daníelsson and de Vries (2001).

8.4 Prediction of M4 Processes

Accurate assessment of financial risks is of great importance for both investors and regulators. A challenging problem is to estimate measures of risk for the cumulative loss over a multiple-day period of a portfolio containing D assets. In particular, we are interested in estimating VaR and ES over a period of $H = 10$ days. For instance, the New Basel Capital Accord specifies that financial institutions should report VaR over a 10-day period.

In this section we propose a simulation-based method based on M4 processes to obtain the joint predictive distribution of future losses over a multiple-day period. Measures of risk (VaR and ES) can then be obtained directly from this simulated distribution.

Previous work on modeling financial data with M4 processes include Zhang and Smith (2001), Zhang (2002) and Smith (2003). However, these works do not discuss the estimation of measures of risk.

8.4.1 Model Description

Let $\{\boldsymbol{r}_t = (r_{t,1}, \ldots, r_{t,D})', \ t = 1, \ldots, T\}$ be daily returns on D assets. The portfolio return from time $t-1$ to time t is given by (8.4), i.e.,

$$r_{t,p} \approx \sum_{d=1}^{D} w_{t,d} r_{t,d}, \tag{8.28}$$

where $w_{t,d}$ is the proportion of wealth invested in asset d at time t. Recalling that log returns are time-additive, the cumulative portfolio return over an horizon of H days is

$$\begin{aligned} r_{t,p}(H) &= r_{t+1,p} + r_{t+2,p} + \cdots + r_{t+H,p} \\ &= \sum_{h=1}^{H} r_{t+h,p}. \end{aligned} \tag{8.29}$$

The relevant question for obtaining multiple-day period measures of risk is

What is the distribution of the cumulative loss between day t and $t+H$?

We consider this question in detail in §8.4.4.

Our strategy to analyze returns is a three-stage approach, similar to McNeil and Frey (2000), Zhang (2002) and Smith (2003). First fit a volatility model to remove the mean and volatility effects. Second, fit an extreme value model to the standardized residuals from the first stage and transform to unit Fréchet scale. Third, fit an M4 process model to the variables in the unit Fréchet scale.

Our model takes into account several empirical facts of financial time series, such as time-varying volatility, heavy tails, and extremal dependence across assets.

Let r_t be the conditional return given past information $\mathcal{F}_t$, then assume

$$r_t = m_t + s_t Y_t. \tag{8.30}$$

It is usual to assume an AR(1) model for the conditional mean m_t, of the form

$$m_t = \phi_0 + \phi_1 m_{t-1}, \tag{8.31}$$

where $\phi_0 \in \mathbb{R}$, $|\phi_1| < 1$. The volatility effect can be modeled with a GARCH(1,1) model,

$$s_t^2 = \beta_0 + \beta_1 \epsilon_{t-1}^2 + \beta_2 s_{t-1}^2, \tag{8.32}$$

where $\epsilon_t = r_t - m_t$, and $\beta_0 > 0$, $\beta_1 \geq 0$, and $\beta_2 \geq 0$.

The first stage is basically the same as in previous works. The idea is to estimate m_t and s_t using a quasi-maximum likelihood approach and then work with the standardized residuals

$$Y_t = \frac{r_t - \hat{m}_t}{\hat{s}_t}. \tag{8.33}$$

However, unlike conventional GARCH models, in our approach we do not assume the standardized residuals are iid—there is dependence in the tails that is not captured by the GARCH model and it will be modeled in the third stage of the procedure.

The second stage is to fit GPD models to the exceedances of Y_t over a large threshold u. Let $\lambda = P\{Y_t > u\}$, and let β and ξ be the scale and shape parameters of the GPD given by (2.9). Then we have the approximation

$$\begin{aligned} 1 - F(y+u) &= P\{Y_t > y+u \mid Y_t > u\} P\{Y_t > u\} \\ &\approx \lambda \left(1 + \xi \frac{y}{\beta}\right)_+^{-1/\xi}. \end{aligned} \tag{8.34}$$

The marginal transformation to the unit Fréchet scale is carried out as follows. If Y has the distribution function F given by (8.34), the transformation

$$X = \left[-\log \left\{ 1 - \lambda \left(1 + \xi \frac{Y-u}{\beta}\right)_+^{-1/\xi} \right\} \right]^{-1} \tag{8.35}$$

induces a random variable X that satisfies

$$P\{X \leq x\} = \exp\{-1/x\}, \qquad x \geq -1/\log(1-\lambda). \tag{8.36}$$

After applying the marginal transformation to the exceedances of each of the D series we obtain a process $\{\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,D})',\ t = 1, \ldots, T\}$, such that each $\{X_t\}$ has unit

Fréchet marginals.

The third stage is to assume that the process $\{\boldsymbol{X}_t\}$ can be approximated by an M4 process

$$X_{td} = \max_{\ell} \max_{k} \alpha_{\ell k d} Z_{\ell,t-k}, \qquad t \in \mathbb{N}, \quad 1 \leq d \leq D. \tag{8.37}$$

The parameters of this model are then estimated using the clustering based method described in Algorithm 3.6.6. This allows us to model the clustering of extreme events (temporal dependence) as well as joint dependence in the extremes across different assets.

8.4.2 One-Step Prediction

In this section we assume that all model parameters have been estimated, so to simplify the exposition we assume the parameters are known.

The prediction of risk is based on the predictive distributions of future losses. Our prediction methods are based on a state-space model representation of M4 processes, given by

$$X_{td} = \max_{\ell} \max_{k} \alpha_{\ell k d} Z_{\ell,t-k}, \qquad t \in \mathbb{N}, \quad 1 \leq d \leq D, \tag{8.38}$$

$$Y_{td} = \widetilde{X}_{td} + \epsilon_{td}, \qquad t \in \mathbb{N}, \quad 1 \leq d \leq D, \tag{8.39}$$

where

$$\widetilde{X} = \mu + \psi \frac{X^{\xi} - 1}{\xi} \tag{8.40}$$

is the marginal transformation from the unit Fréchet to the GEV scale.

Our approach is as follows. Given the parameters of the model, we obtain the filtering distributions of the states $\{Z_{\ell t}\}$ using particle filters. Then we use the filtering distribution to simulate the predictive distributions of future states, which in turn is used to obtain the predictive distribution of future losses. This approach is used for both one-step and multiple-step predictions.

The one-step prediction equation is

$$f\left(x_{t+1} \mid y_{1:t}\right) = \int f\left(x_{t+1} \mid x_{0:t}\right) f\left(x_{0:t} \mid y_{1:t}\right) \mathrm{d}x_{0:t}. \tag{8.41}$$

It is not possible to compute this expression explicitly, however a Monte Carlo sample from this distribution can be obtained with the following procedure:

1. Run a particle filter to obtain a sample from the filtering distribution of the last K states, i.e., $\left\{Z^{(i)}_{\ell,t-K+1:t}\right\}$, $1 \le \ell \le L$, $1 \le i \le N$.

2. Generate innovations $Z^{(i)}_{\ell,t+1} \sim \Phi_1(x)$, $1 \le \ell \le L$, $1 \le i \le N$.

3. Given α and $\left\{Z^{(i)}_{\ell,t-K+1:t+1}\right\}$, obtain $\boldsymbol{X}^{(i)}_{t+1}$, $1 \le i \le N$ from (8.38).

Then $\boldsymbol{X}^{(i)}_{t+1}$, $1 \le i \le N$ can be considered as a sample from $f\left(x_{t+1} \mid y_{1:t}\right)$.

8.4.3 Multiple-Step Prediction

To obtain the two-step predictive distribution consider the joint prediction

$$f\left(x_{t+1}, x_{t+2} \mid y_{1:t}\right) = f\left(x_{t+2} \mid x_{t+1}\right) f\left(x_{t+1} \mid y_{1:t}\right). \tag{8.42}$$

In general, the joint H-step predictive distribution is

$$f\left(x_{t+1:t+H} \mid y_{1:t}\right) = \prod_{h=1}^{H-1} f\left(x_{t+h+1} \mid x_{t+h}\right) f\left(x_{t+1} \mid y_{1:t}\right). \tag{8.43}$$

Our approach to simulate the multiple-step predictive distribution is a straightforward extension of the methods for one-step predictions. In this case generate innovations

$$Z^{(i)}_{\ell,t+h} \sim \Phi_1(x), \qquad 1 \le \ell \le L, \quad 1 \le h \le H, \quad 1 \le i \le N. \tag{8.44}$$

Given these innovations and the particle filtering sample $\left\{Z^{(i)}_{\ell,t-K+1:t}\right\}$, use (8.38) to obtain sample paths

$$\boldsymbol{X}^{(i)}_{t+1}, \ldots, \boldsymbol{X}^{(i)}_{t+H}, \qquad 1 \le i \le N. \tag{8.45}$$

8.4.4 Risk Assessment Based on M4 Processes

In this section we address the question

What is the distribution of the cumulative loss between day t and $t + H$?

More precisely, we need to analyze the predictive distribution of the cumulative portfolio return.

We propose to use predictions from an M4 process to approximate the joint extremal behavior of future losses. However, since the goal is to obtain a sample from the distribution of cumulative returns, we also need to account for the center of the distribution. To achieve this we adapt an approach previously proposed by McNeil and Frey (2000) and Daníelsson and de Vries (2001).

The idea is to fit two independent M4 process $\left\{\boldsymbol{x}_t^{(\mathsf{L})}\right\}$ and $\left\{\boldsymbol{x}_t^{(\mathsf{P})}\right\}$, for the lower tail (losses) and upper tail (profits), respectively. Then obtain bootstrap samples from the center of the standardized residuals for the center of the distribution. The steps for this procedure are as follows.

1. Generate a sample path $\boldsymbol{x}_{t+1}^{(\mathsf{L})}, \ldots, \boldsymbol{x}_{t+H}^{(\mathsf{L})}$ from the fitted M4 process for the lower tail (losses) and transform to the GEV scale, i.e., $\widetilde{\boldsymbol{x}}_{t+1}^{(\mathsf{L})}, \ldots, \widetilde{\boldsymbol{x}}_{t+H}^{(\mathsf{L})}$.

2. Generate a sample path $\boldsymbol{x}_{t+1}^{(\mathsf{P})}, \ldots, \boldsymbol{x}_{t+H}^{(\mathsf{P})}$ from the fitted M4 process for the upper tail (profits) and transform to the GEV scale, i.e., $\widetilde{\boldsymbol{x}}_{t+1}^{(\mathsf{P})}, \ldots, \widetilde{\boldsymbol{x}}_{t+H}^{(\mathsf{P})}$.

3. Obtain a bootstrap sample $\boldsymbol{e}_{t+1}, \ldots, \boldsymbol{e}_{t+H}$, where each $e_{t+h,d}$, $1 \leq h \leq H$, $1 \leq d \leq D$ is one of the standardized residuals (8.33).

4. For $h = 1, \ldots, H$:

   - set $\boldsymbol{Y}_{t+h} = \boldsymbol{e}_{t+h}$;
   - if $\tilde{x}_{t+h,d}^{(\mathsf{L})} > u^{(\mathsf{L})}$ set $Y_{t+h,d} = \tilde{x}_{t+h,d}^{(\mathsf{L})}$, $d = 1, \ldots, D$;
   - if $\tilde{x}_{t+h,d}^{(\mathsf{P})} < -u^{(\mathsf{P})}$ set $Y_{t+h,d} = \tilde{x}_{t+h,d}^{(\mathsf{P})}$, $d = 1, \ldots, D$.

We simulate N sample paths from the procedure described above, i.e.,

$$\boldsymbol{Y}_{t+1}^{(i)}, \boldsymbol{Y}_{t+2}^{(i)}, \ldots, \boldsymbol{Y}_{t+H}^{(i)}, \qquad 1 \leq i \leq N, \tag{8.46}$$

which combined with mean and volatility predictions

$$\hat{m}_{t+h,d}, \ \hat{s}_{t+h,d}, \qquad 1 \leq h \leq H, \quad 1 \leq d \leq D, \tag{8.47}$$

give N simulated paths of returns

$$r_{t+h,d}^{(i)} = \hat{m}_{t+h,d} + \hat{s}_{t+h,d} Y_{t+h,d}^{(i)}, \qquad 1 \le h \le H, \quad 1 \le d \le D, \quad 1 \le i \le N. \tag{8.48}$$

Finally, for each simulated path, we calculate the approximate portfolio return by (8.28), and the cumulative return by (8.29). Measures of risk can then be obtained directly from this simulated distribution.

8.5 Data Analysis of Nasdaq Sector Indices

In this section we analyze four Nasdaq sector indices[1]: Nasdaq Bank (BK), Nasdaq Industrial (ID), Nasdaq Insurance (IS), and Nasdaq Transportation (TR). The data are 5848 daily negative log returns from January 4, 1982 to March 18, 2005. A plot of negative returns calculated from the data is given in Figure 8.1. Percentiles of the negative returns are shown in Table 8.1.

Our goal is to obtain multiple-day measures of risk for a hypothetical portfolio of these assets. In particular we are interested in the distribution of the cumulative loss over a 10-day period. We follow the approach described in the preceding section: First, we simulate sample paths of the predictive distribution of the multivariate time series based on M4 processes fitted to the tails. Then we combine the multiple simulated paths to obtain a sample of the predictive distribution of the cumulative return of the portfolio up to time $T + 10$. Measures of risk can be obtained directly from this simulated distribution.

8.5.1 Mean and Volatility Models

In the first stage of modeling we fit AR(1)-GARCH(1,1) models defined by (8.30)–(8.32) to each of the individual series in order to remove trend and volatility effects. Then we calculate standardized residuals according to (8.33), which are shown in Figure 8.2.

The choice of GARCH-type models can be replaced by other volatility model, such as a MIDAS model. We believe that the choice of volatility model does not have any qualitative effect on later modeling stages, however a proper comparison and further research is nec-

[1]Refer to `http://dynamic.nasdaq.com/reference/IndexDescriptions.stm` for a description of the data. The data source is `http://finance.yahoo.com`.

essary in this respect. We choose to fit GARCH models for convenience because software is readily available in the S+FinMetrics module of S-PLUS.

The top part of Table 8.2 shows the correlation between standardized residuals from GARCH models. Linear dependence is weak as the correlation between variables ranges from 0.5 (BK vs. TR) to 0.67 (BK vs. ID). The bottom part of Table 8.2 shows correlation for standardized residuals where at least one of the variables exceeds the threshold $u = 1.45$. Linear dependence is very weak, the smallest being BK vs. TR and the strongest ID vs TR with an estimate of 0.43. Even though there is some linear dependence that we are not taking into account in our model, this dependence is rather small in the tails.

8.5.2 Empirical Evidence for M4 Processes

We obtain estimates of the extremal index of the standardized residuals using the runs method (Smith and Weissman, 1994). We choose $r = 2$ consecutive observations below the threshold $u = 1.45$ as separating two clusters of exceedances. Table 8.3 shows that these estimates fluctuate around 0.8068 and 0.9156, so the mean cluster size of exceedances is in the interval $(1.10, 1.24)$, which indicates that there is clustering of extremal events that is not captured by the volatility model. However this time-dependence is not very strong.

Plots of the exceedances over the threshold of the standardized residuals are shown in Figure 8.4 (losses) and Figure 8.5 (profits). These plots show that there are joint exceedances over the threshold, suggesting joint extremal dependence across series.

The next step is to fit univariate extreme value models to each of the series in order to transform the marginals to the unit Fréchet scale. We fit models for both tails (losses and profits) of the standardized residuals from GARCH models. Figure 8.3 shows mean excess plots with approximate 95% confidence intervals for each of the variables. These plots are reasonably linear above the threshold $u = 1.45$, which suggests that the data can be modeled with a GPD above this threshold. Hence we select a threshold of $u = 1.45$ for all variables.

Parameter estimates of the point process models based on the threshold $u = 1.45$ and 252 observations per year are shown in Table 8.4. Since there are usually 252 trading days in a year, these estimates correspond to the parameters of the limiting distribution of

annual maxima of each series.

Table 8.5 shows the parameters of the GPD models fitted to the data. As expected the estimates of the shape parameter ξ are the same for the point process and the GPD models. In what follows, we use the GPD parameter estimates because the GPD parameterization is more convenient for daily data, and also because the two sets of parameters (λ, β, ξ) and (μ, ψ, ξ) are consistent above the threshold, as shown by (2.12) in Chapter 2. Diagnostic plots for each of the GPD models are shown in Figures 8.6 and 8.7. Each column shows diagnostic plots for one variable, where the diagnostics correspond to a plot of the excess distribution (first row), a plot of the tail of underlying distribution (second row), a scatter plot of residuals (third row), and a Q-Q plot of residuals (fourth row). These plots indicate good fits for all 8 variables.

We use the GPD parameter estimates for transforming the exceedances over the threshold $u = 1.45$ to the unit Fréchet scale. Note that the estimate of $\lambda_d,\ d = 1, 2, 3, 4$ is different for each of the variables, so the transformed variables are exceedances over a threshold $-1/\log(\lambda_d),\ d = 1, 2, 3, 4$ in the unit Fréchet scale shown in Table 8.6.

Figures 8.8–8.11 show scatter plots of joint exceedances. The numbers at the top-right corner show the number of expected joint exceedances under the assumption of independence, and the actual number of observed exceedances, which is the number of points in the plot.

Figure 8.8 shows scatter plots of joint exceedances at 1, 2, and 3 lags of the same series. The difference between expected and observed exceedances do not suggest strong dependence in time.

Figure 8.9 shows scatter plots of joint exceedances across variables in the same day. The lower diagonal corresponds to losses, and the upper diagonal corresponds to profits. The large differences between expected and observed exceedances do suggest strong evidence of dependence across series.

Figure 8.10 shows scatter plots of joint exceedances across variables at different lags. There is no strong evidence of lag dependence across series.

We are also interested to find if there is extremal dependence between the upper and lower tails. The question is based on the observation that if there is a large event on

any given day, then it may be likely to observe an extreme event in the same or opposite direction during the following days. Figure 8.11 shows scatter plots of joint exceedances of the upper and lower tails of individual series at 1 and 2 lags. Again, there is not much difference between expected and observed exceedances, except for the ID variable.

8.5.3 Estimation of M4 Processes

We specify an M4 process model in 4 dimensions with $L = 10$ signature patterns and 5 lags $|k| \leq 2$, and we fit this model using the clustering-based method described in Algorithm 3.6.6. Table 8.7 shows the parameter estimates $\{\alpha_{\ell kd}\}$ of the M4 process model fitted to the upper tail of the Nasdaq Indices data (losses). Table 8.8 shows the within-cluster sum of squares and the cluster size for each signature pattern identified by K-means. We also fit a model for the lower tail (losses) with the same number of lags and signature patterns, but omit the corresponding tables. Figures 8.12–8.13 show plots of the parameter estimates of these two fitted M4 process models.

An alternative model is to fit an M4 process in 8 dimensions that accounts for extremal dependence across the two tails. Even though there is not much empirical evidence that suggest there is extremal dependence across tails, we fit this model to illustrate its main features. We choose $L = 5$ signature patterns and 5 lags $|k| \leq 2$. Figure 8.14 shows the parameter estimates for this model, where each signature pattern is displayed with a different line pattern. For instance, consider the signature pattern that is displayed with a solid line. In this case the parameters corresponding to losses have an A-shaped form, while the parameters corresponding to profits have a V-shaped form. The opposite shape of parameters for losses and profits is consistent across all signature patterns.

The estimated M4 parameters (Table 8.7) are used to obtain estimates of the extremal indices of the individual series and of the multivariate extremal index by a direct application of (3.13). The results are shown in Table 8.9. It is important to remember that the univariate extremal index does not depend on the threshold chosen, but the multivariate extremal index does. The estimate of the multivariate extremal index is evaluated at $\boldsymbol{\tau} = (1, 1, 1, 1)$, however we have tried other values of $\boldsymbol{\tau}$ and the estimate is not very sensitive to the choice of τ.

8.5.4 Risk Assessment

We use the data and the parameters of the two fitted M4 processes ($L = 10$, $|k| \leq 2$) to obtain filtering samples of the unobserved states. We select $N = 1000$ so the number of particles is relatively small. Then we simulate $N = 1000$ sample paths of the predictive distribution over a period of 10 days as described in §8.4, and obtain the corresponding VaR and ES. Our results are shown in Table 8.10. We present the individual and hypothetical portfolio risks at the 95%, 97.5%, 99%, 99.5%, and 99.9% levels.

8.5.5 Backtesting

We perform backtesting to assess the effectiveness of our methods in predicting future losses at 10-day horizons. The idea is to obtain in-sample estimates of VaR and ES over a 10-day period as described in §8.4. Then compare them with the cumulative loss that was actually observed over the same period. If the VaR and ES estimated for that period is smaller than the actual value then we say there has been a violation.

We consider blocks of 3839 observations at a time, and for each block obtain measures of risk using $N = 1000$, consequently we end up with a total of $n = 2000$ comparisons. Under the assumption that our method estimates the quantile q without bias, the expected number of violations is $(1 - q)n$ where n is the number of backtest experiments. For the ES we do not have a benchmark for expected violations. For $n = 2000$, and levels of 0.95, 0.975, 0.99, 0.995, and 0.999, the expected number of violations are 100, 50, 20, 10, 2, respectively.

The number of VaR violations is presented in Table 8.11. For comparison we fit three different types of models. First, a model where events in the tails are generated as iid observations from a GPD distribution, similar to the method described in McNeil and Frey (2000). Second, a model where separate M4 process are fitted to the upper and lower tails, i.e., two different four-dimensional processes. Third, a joint M4 process model for the upper and lower tails, i.e., a single process in eight dimensions. These models are labeled in Table 8.11 as "iid", "M4 (4-D)", and "M4 (8-D)", respectively. The main observations from these tables are as follows.

The risk of BK is overestimated at any level of risk by all of the methods. Similarly,

the risk of ID is overestimated at levels of 97.5% and 99% by all of the methods. However, the M4 process models seem to estimate the risk of ID without bias at levels of 99.5% and 99.9%. The risk of IS is correctly estimated by all of the methods. The risk of TR is underestimated by the iid method, especially for levels above 99%; the M4 processes perform better for the TR variable because the violations of M4 process models are much closer to the expected number, even for very high quantiles. On the other hand, the portfolio risk is underestimated by all of the methods. However, the violations with the iid method are always more than any of the M4 process models, suggesting that there is an improvement by modeling the joint extremes with M4 processes.

The corresponding results for the ES are shown in Table 8.12. In this case we do not have a benchmark for expected violations, but the results are consistent with the results for VaR. In particular, the violations of the portfolio ES from the iid method are always more than those from M4 processes.

8.5.6 Conclusions

We have successfully applied the methods developed in this work to estimating measures of risk of financial portfolios.

Risk assessment based on M4 processes gives the advantage of modeling clustering of extremes and extremal dependence across variables, which are features empirically observed in financial data, but not captured by ARCH-type models.

M4 process models are promising models for portfolios of many variables, which are typical in financial applications. In this work we consider a portfolio of four variables, but there is no restriction on the number of assets that can be included in the analysis.

Even though we have made progress in fitting these models to real data, there are a number of alternative analyses and comparisons that need to be explored.

Percentile	Bank	Industrial	Insurance	Transport.
0%	−0.0807	−0.1052	−0.0536	−0.0852
1%	−0.0210	−0.0352	−0.0234	−0.0279
2.5%	−0.0153	−0.0251	−0.0173	−0.0213
5%	−0.0119	−0.0187	−0.0133	−0.0169
8%	−0.0093	−0.0147	−0.0108	−0.0136
10%	−0.0083	−0.0130	−0.0097	−0.0123
25%	−0.0042	−0.0067	−0.0051	−0.0063
50%	−0.0008	−0.0010	−0.0007	−0.0004
75%	0.0027	0.0051	0.0039	0.0051
90%	0.0078	0.0140	0.0090	0.0115
92%	0.0091	0.0159	0.0104	0.0130
95%	0.0117	0.0204	0.0131	0.0163
97.5%	0.0161	0.0279	0.0174	0.0214
99%	0.0236	0.0379	0.0240	0.0287
99.9%	0.0417	0.0787	0.0410	0.0507
100%	0.0922	0.1419	0.1000	0.1399

Table 8.1: *Percentiles of Negative Returns of Nasdaq Indices.*

Bank	Industrial	Insurance	Transport.
All Observations			
1.0000	0.5854	0.5805	0.5005
0.5854	1.0000	0.6266	0.6722
0.5805	0.6266	1.0000	0.5335
0.5005	0.6722	0.5335	1.0000
At Least One Exceedance			
1.0000	0.2503	0.3243	0.1563
0.2503	1.0000	0.3126	0.4292
0.3243	0.3126	1.0000	0.1903
0.1563	0.4292	0.1903	1.0000

Table 8.2: *Correlation Between Standardized Residuals. Top part shows correlation coefficients computed using all observations. Bottom part shows correlation coefficients computed using only rows of observations where at least one of the variables exceeds the threshold* $u = 1.45$.

	Bank	Industrial	Insurance	Transport.
Losses	0.8368	0.8068	0.8241	0.8627
Profits	0.9070	0.9156	0.8506	0.8827

Table 8.3: *Estimates of the Extremal Index of Standardized Residuals. Estimates obtained with the runs method, where runs are defined whenever* $r = 2$ *consecutive observations fall below the threshold of* $u = 1.45$*.*

Series	Number of Exceedances	μ (s.e.)	ψ (s.e.)	ξ (s.e.)
Losses				
Bank	386	3.6423	0.9180	0.1195
		(0.1528)	(0.1168)	(0.0597)
Industrial	471	3.4846	0.8066	0.1212
		(0.1307)	(0.0910)	(0.0469)
Insurance	432	3.3887	0.7791	0.1135
		(0.1285)	(0.0939)	(0.0533)
Transport.	386	3.3270	0.8175	0.1493
		(0.1333)	(0.0967)	(0.0521)
Profits				
Bank	344	2.9017	0.5701	0.0429
		(0.0929)	(0.0613)	(0.0526)
Industrial	308	2.5555	0.4126	−0.0273
		(0.0670)	(0.0398)	(0.0508)
Insurance	348	2.7960	0.4922	−0.0073
		(0.0809)	(0.0551)	(0.0579)
Transport.	358	2.7977	0.5238	0.0455
		(0.0856)	(0.0579)	(0.0538)

Table 8.4: *Parameter Estimates of Poisson Process Models. Models Fitted to Nasdaq Indices Standardized Residuals. Threshold is* $u = 1.45$*, and number of observations per year are* 252*. Standard errors in parenthesis.*

Series	Number of Exceedances	λ	β (s.e.)	ξ (s.e.)
Losses				
Bank	386	0.06601	0.6562	0.1195
			(0.0514)	(0.0599)
Industrial	471	0.08054	0.5602	0.1209
			(0.0367)	(0.0469)
Insurance	432	0.07387	0.5592	0.1133
			(0.0401)	(0.0533)
Transport.	386	0.06601	0.5373	0.1492
			(0.0389)	(0.0521)
Profits				
Bank	344	0.05882	0.5080	0.0426
			(0.0383)	(0.0526)
Industrial	308	0.05267	0.4428	−0.0273
			(0.0338)	(0.0508)
Insurance	348	0.05951	0.5019	−0.0071
			(0.0396)	(0.0579)
Transport.	358	0.06122	0.4625	0.0456
			(0.0349)	(0.0538)

Table 8.5: *Parameter Estimates of GPD Models. Models Fitted to Nasdaq Indices Standardized Residuals. Threshold is* $u = 1.45$. *Standard errors in parenthesis.*

	Bank	Industrial	Insurance	Transport.
Losses	14.6446	11.9091	13.0306	14.6446
Profits	16.4949	18.4825	16.2995	15.8299

Table 8.6: *Thresholds in Unit Fréchet Scale.*

ℓ	$\alpha_{\ell,-2}$	$\alpha_{\ell,-1}$	$\alpha_{\ell,0}$	$\alpha_{\ell,1}$	$\alpha_{\ell,2}$
Bank					
1	0.0000	0.0001	0.0078	0.0004	0.0011
2	0.0012	0.0000	0.1218	0.0002	0.0000
3	0.0011	0.0057	0.0178	0.0060	0.0008
4	0.0000	0.0000	0.4410	0.0010	0.0007
5	0.0082	0.0021	0.0231	0.0038	0.0038
6	0.0030	0.0021	0.0087	0.0133	0.0003
7	0.0047	0.0004	0.0185	0.0000	0.0016
8	0.0000	0.0008	0.1065	0.0000	0.0000
9	0.0065	0.0080	0.0147	0.0083	0.0042
10	0.0018	0.0108	0.1194	0.0116	0.0067
Industrial					
1	0.0003	0.0003	0.0089	0.0000	0.0006
2	0.0007	0.0000	0.0462	0.0005	0.0000
3	0.0039	0.0020	0.1382	0.0028	0.0040
4	0.0004	0.0009	0.0462	0.0000	0.0012
5	0.0017	0.0043	0.4502	0.0000	0.0071
6	0.0102	0.0018	0.0066	0.0154	0.0084
7	0.0021	0.0001	0.0970	0.0000	0.0023
8	0.0006	0.0000	0.0454	0.0000	0.0000
9	0.0041	0.0118	0.0213	0.0049	0.0095
10	0.0055	0.0063	0.0069	0.0088	0.0106
Insurance					
1	0.0010	0.0009	0.3871	0.0004	0.0000
2	0.0027	0.0003	0.1252	0.0004	0.0000
3	0.0000	0.0037	0.0188	0.0045	0.0016
4	0.0019	0.0014	0.0132	0.0010	0.0003
5	0.0057	0.0048	0.0207	0.0092	0.0088
6	0.0034	0.0053	0.0971	0.0017	0.0060
7	0.0011	0.0000	0.1173	0.0000	0.0000
8	0.0015	0.0004	0.0455	0.0000	0.0015
9	0.0063	0.0039	0.0433	0.0009	0.0053
10	0.0047	0.0087	0.0184	0.0061	0.0085
Transport.					
1	0.0013	0.0000	0.0167	0.0004	0.0021
2	0.0003	0.0009	0.0137	0.0006	0.0000
3	0.0030	0.0008	0.1769	0.0063	0.0019
4	0.0006	0.0003	0.0084	0.0002	0.0010
5	0.0074	0.0032	0.0230	0.0027	0.0039
6	0.0044	0.0000	0.0023	0.0119	0.0053
7	0.0003	0.0008	0.0292	0.0000	0.0000
8	0.0000	0.0000	0.1000	0.0000	0.0000
9	0.0051	0.0086	0.5186	0.0003	0.0009
10	0.0085	0.0059	0.0067	0.0035	0.0123

Table 8.7: *Parameter Estimates of M4 Process Model (Losses). Model Fitted to Nasdaq Indices Standardized Residuals with Clustering-Based Method,* $-2 \leq k \leq 2$, $L = 10$.

ℓ	Cluster Size	Within-Cluster SS
1	88	2.6396
2	33	6.6856
3	51	10.0044
4	97	7.7971
5	120	12.2522
6	37	7.6160
7	31	6.3058
8	26	7.1899
9	126	15.8855
10	45	9.3575

Table 8.8: *Within-Cluster Sum of Squares and Cluster Size of Signature Patterns. Model Fitted to Nasdaq Indices Standardized Residuals with Clustering-Based Method,* $-2 \leq k \leq 2$*,* $L = 10$*.*

	Multiv.	Bank	Industrial	Insurance	Transport.
Losses	0.9285	0.8840	0.8795	0.8866	0.9106
Profits	0.9551	0.9265	0.9420	0.9060	0.9159

Table 8.9: *Estimates of the Extremal Index Based on M4 Process Parameters. The first column is the estimate of the multivariate extremal index evaluated at* $\boldsymbol{\tau} = (1, 1, 1, 1)$*.*

Level	Bank	Industrial	Insurance	Transport.	Portfolio
95.0%					
VaR	0.0303	0.0355	0.0319	0.0508	0.0205
ES	0.0436	0.0497	0.0562	0.0687	0.0275
97.5%					
VaR	0.0408	0.0427	0.0435	0.0611	0.0251
ES	0.0523	0.0607	0.0746	0.0817	0.0324
99.0%					
VaR	0.0511	0.0620	0.0548	0.0796	0.0331
ES	0.0624	0.0748	0.1129	0.1006	0.0377
99.5%					
VaR	0.0551	0.0681	0.0900	0.0935	0.0366
ES	0.0710	0.0831	0.1582	0.1116	0.0405
99.9%					
VaR	0.0803	0.0793	0.1754	0.1113	0.0427
ES	0.0989	0.1012	0.1944	0.1444	0.0437

Table 8.10: *Value at Risk and Expected Shortfall over a 10-Day Horizon. Estimates Based on M4 Processes. The portfolio column is the risk of an hypothetical portfolio of these assets.*

VaR Level	Bank	Industrial	Insurance	Transport.	Portfolio
95.0%, Expected: 100					
iid	54	70	91	94	267
M4 (4-D)	60	91	107	110	240
M4 (8-D)	67	81	100	104	222
97.5%, Expected: 50					
iid	23	28	44	53	177
M4 (4-D)	21	31	40	51	148
M4 (8-D)	20	27	39	47	139
99.0%, Expected: 20					
iid	4	13	17	23	116
M4 (4-D)	2	12	13	22	81
M4 (8-D)	3	12	11	16	71
99.5%, Expected: 10					
iid	3	9	9	16	79
M4 (4-D)	1	7	7	9	54
M4 (8-D)	1	8	4	8	58
99.9%, Expected: 2					
iid	1	5	2	5	51
M4 (4-D)	0	2	3	3	31
M4 (8-D)	0	3	1	2	26

Table 8.11: *Number of Value at Risk Violations from Backtesting.* $n = 2000$ *models,* $T = 3839$*,* $N = 1000$

ES Level	Bank	Industrial	Insurance	Transport.	Portfolio
95.0%					
iid	12	21	28	39	154
M4 (4-D)	9	20	28	33	122
M4 (8-D)	11	17	25	31	116
97.5%					
iid	3	13	13	22	110
M4 (4-D)	2	9	12	18	76
M4 (8-D)	3	11	9	14	69
99.0%					
iid	1	8	6	12	68
M4 (4-D)	1	3	5	7	44
M4 (8-D)	1	6	1	6	48
99.5%					
iid	1	6	3	7	51
M4 (4-D)	0	2	3	4	35
M4 (8-D)	0	2	1	3	30
99.9%					
iid	1	2	1	4	38
M4 (4-D)	0	2	0	2	23
M4 (8-D)	0	2	0	1	20

Table 8.12: *Number of Expected Shortfall Violations from Backtesting.* $n = 2000$ *models,* $T = 3839$*,* $N = 1000$

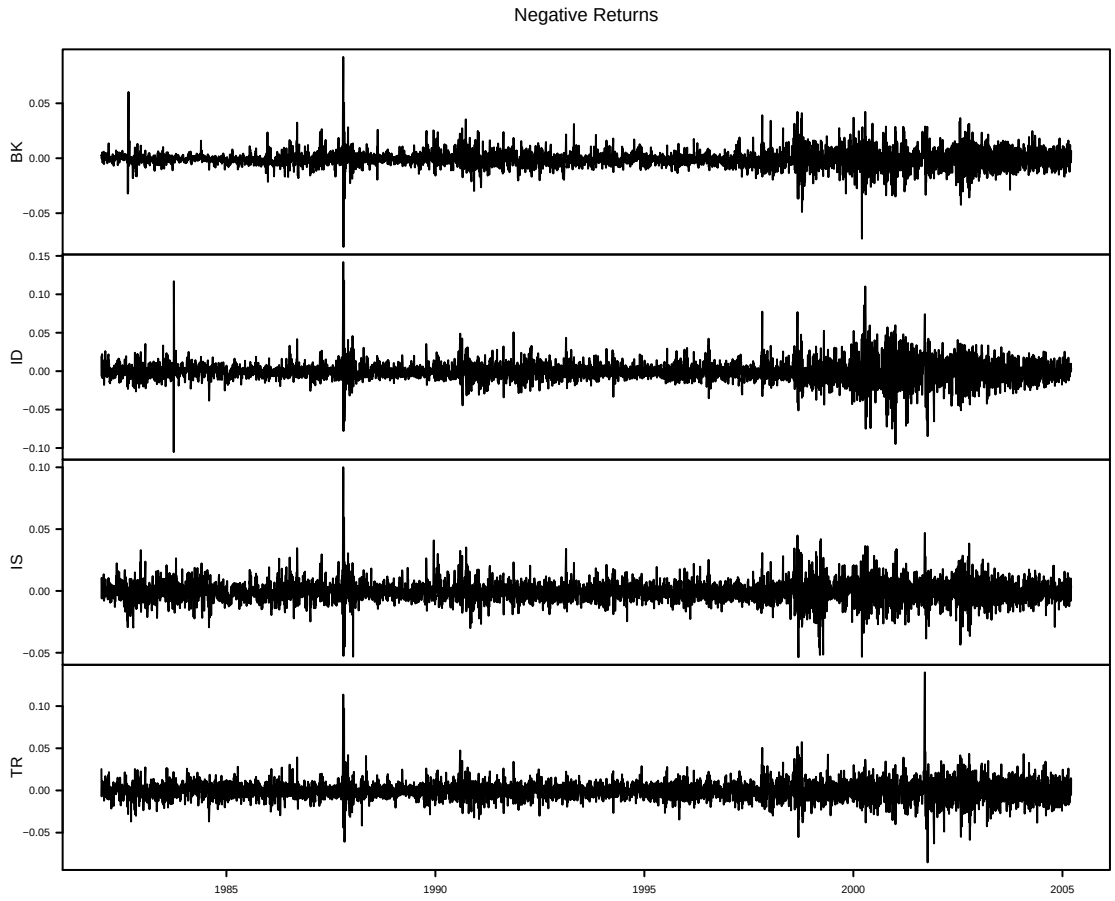

Figure 8.1: *Daily Negative Log Returns of Nasdaq Indices.*

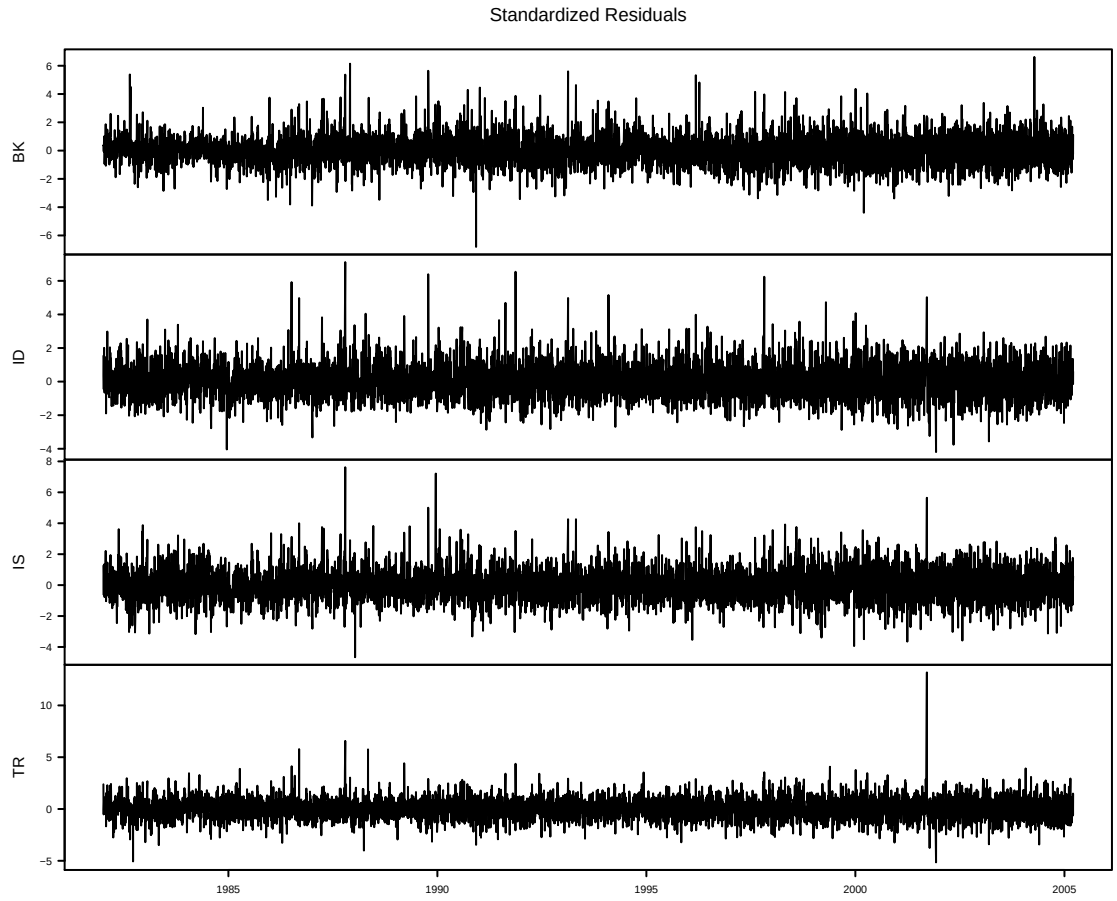

Figure 8.2: *GARCH Standardized Residuals of Nasdaq Indices.*

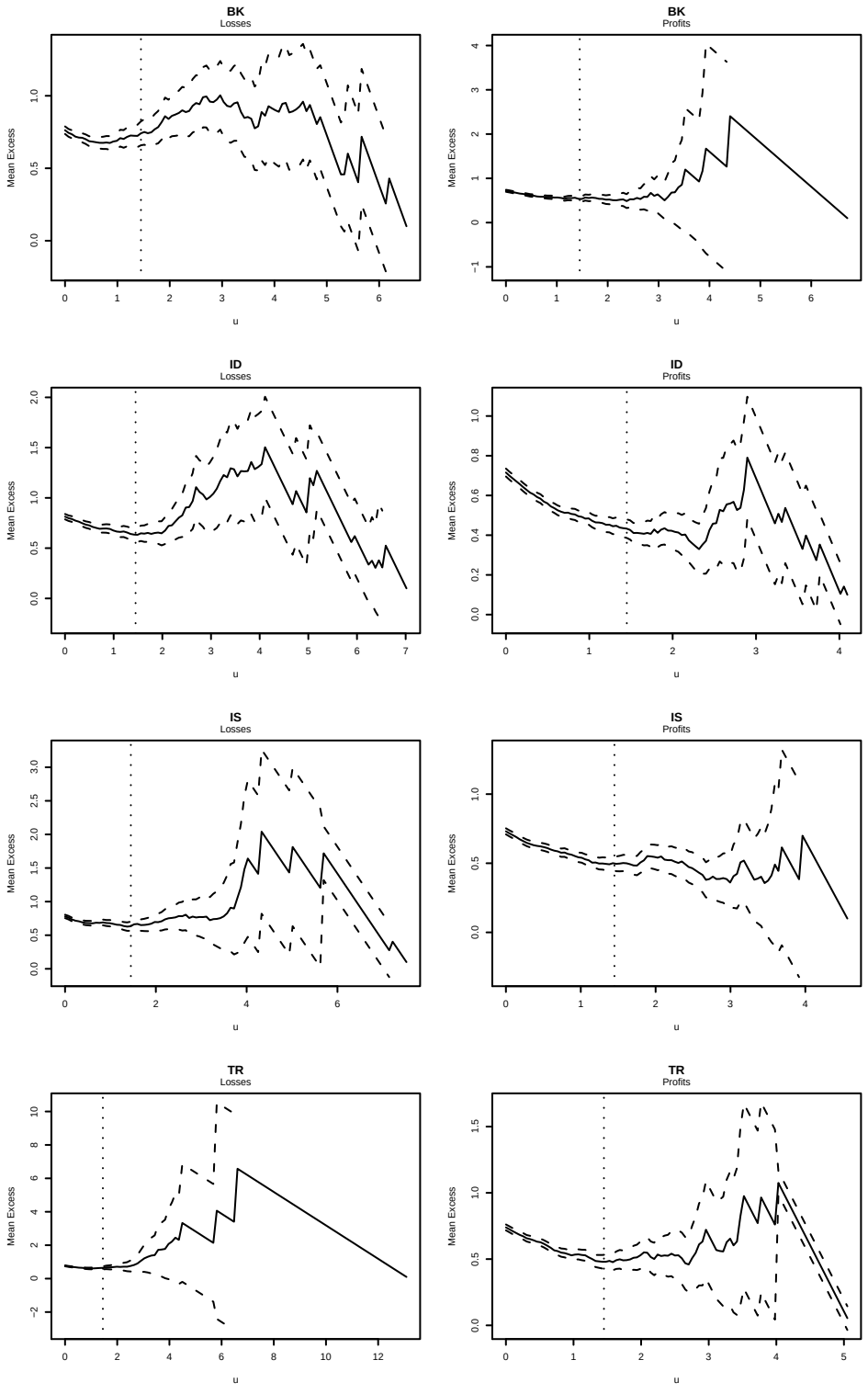

Figure 8.3: *Mean Excess Plots of Nasdaq Indices Standardized Residuals. Reference line at* $u = 1.45$.

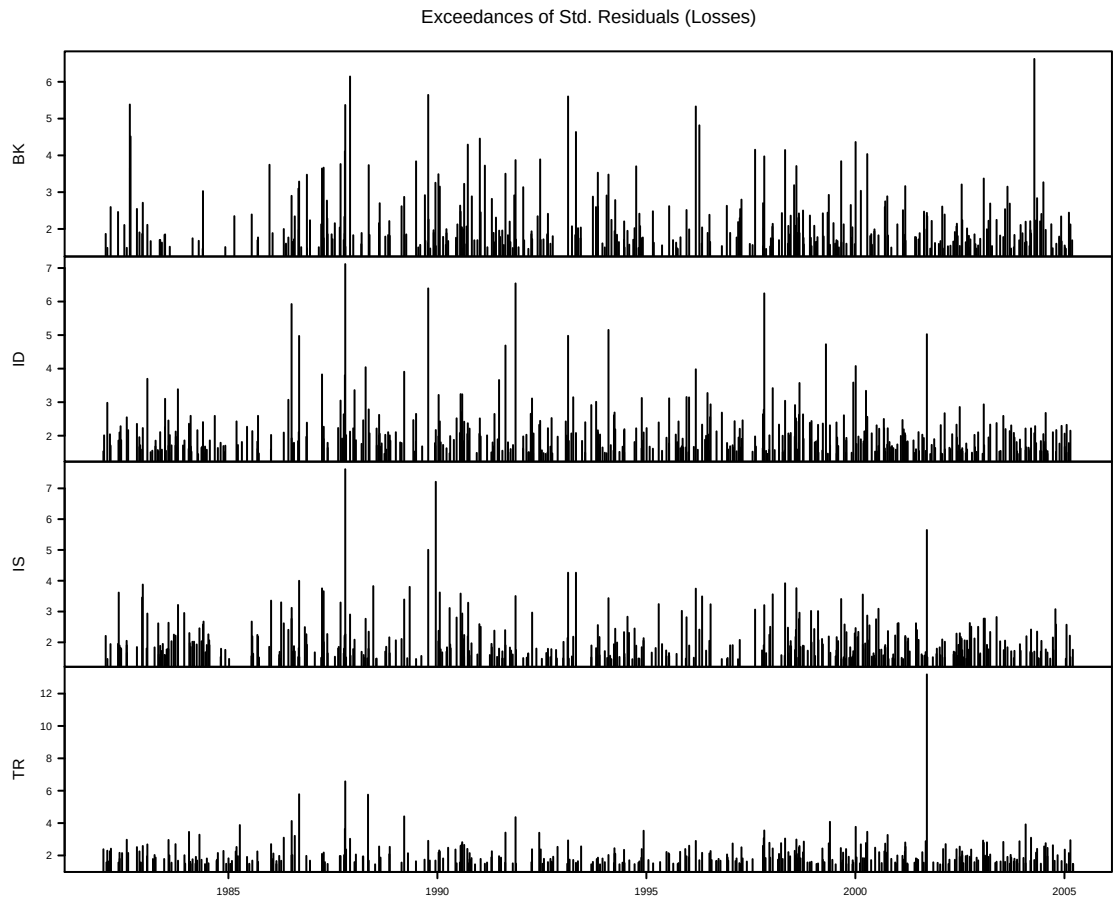

Figure 8.4: *Threshold Exceedances of Standardized Residuals (Losses).*

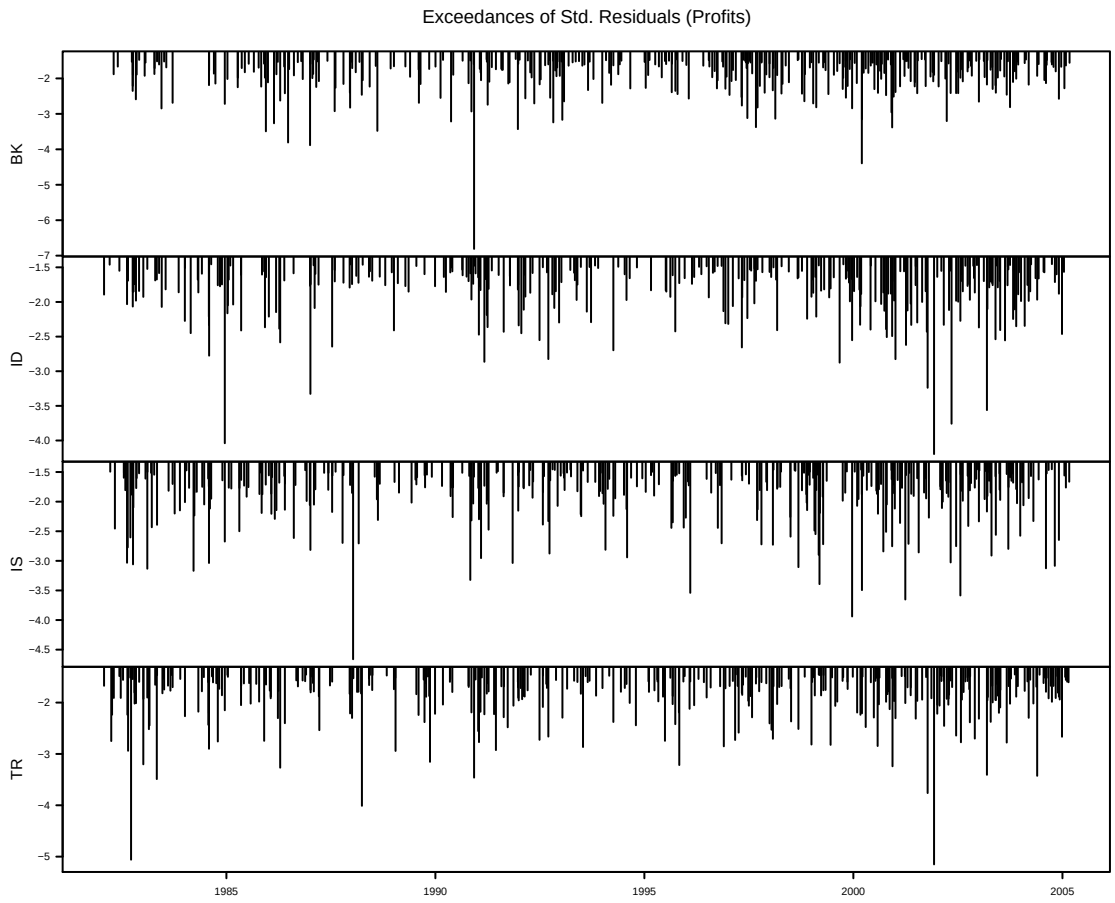

Figure 8.5: *Threshold Exceedances of Standardized Residuals (Profits).*

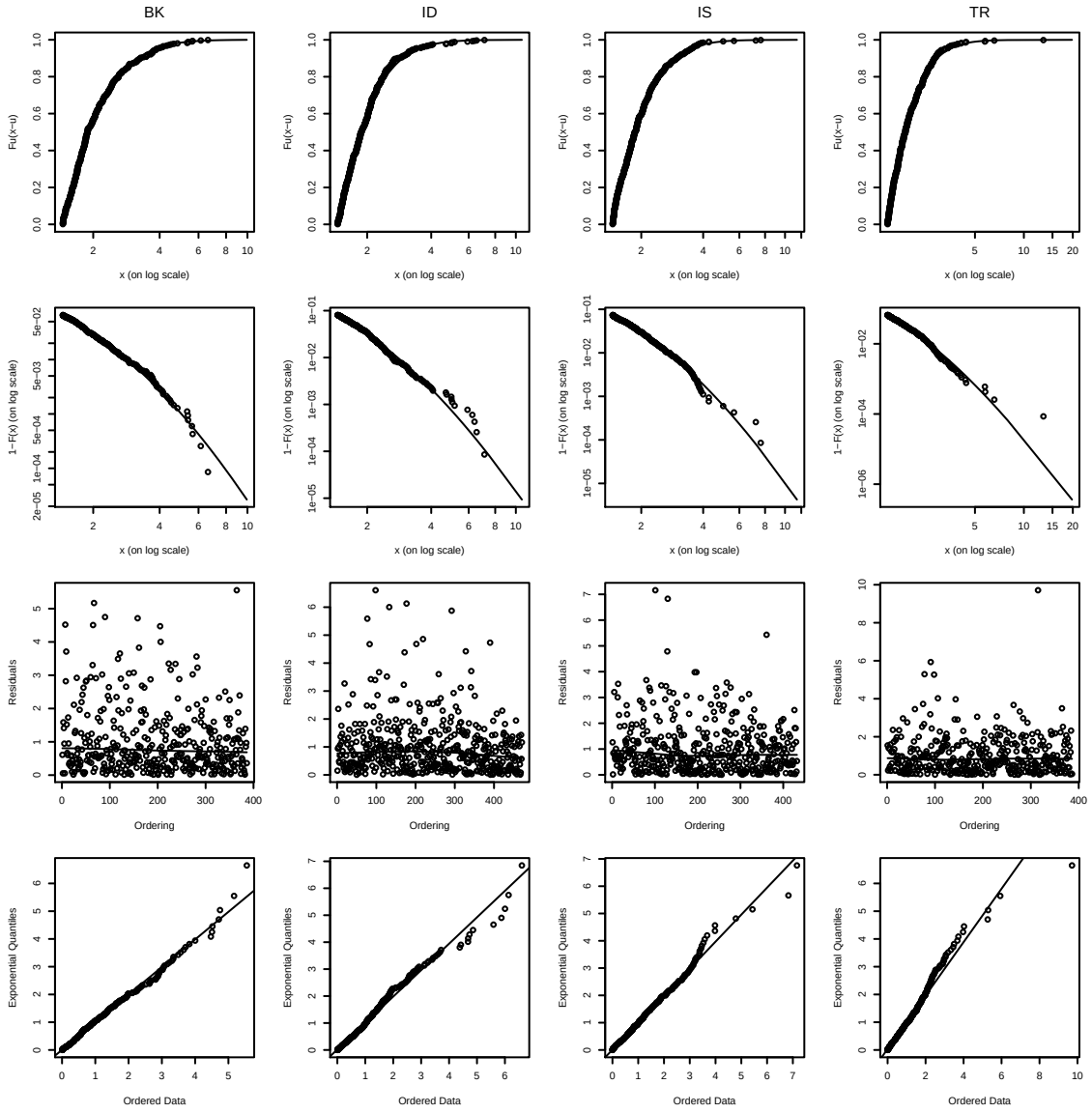

Figure 8.6: *Diagnostic Plots for GPD Models (Losses).*

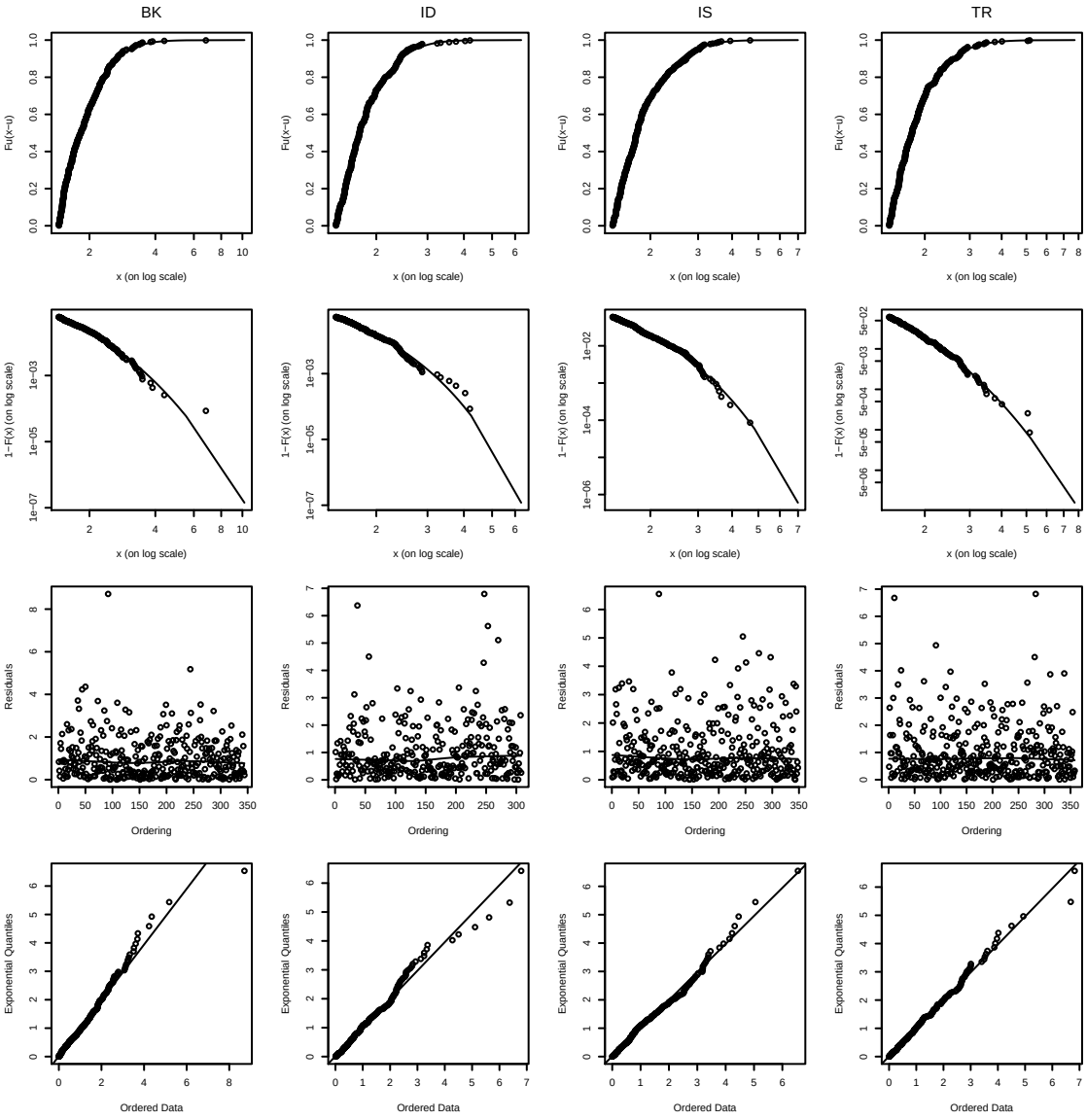

Figure 8.7: *Diagnostic Plots for GPD Models (Profits).*

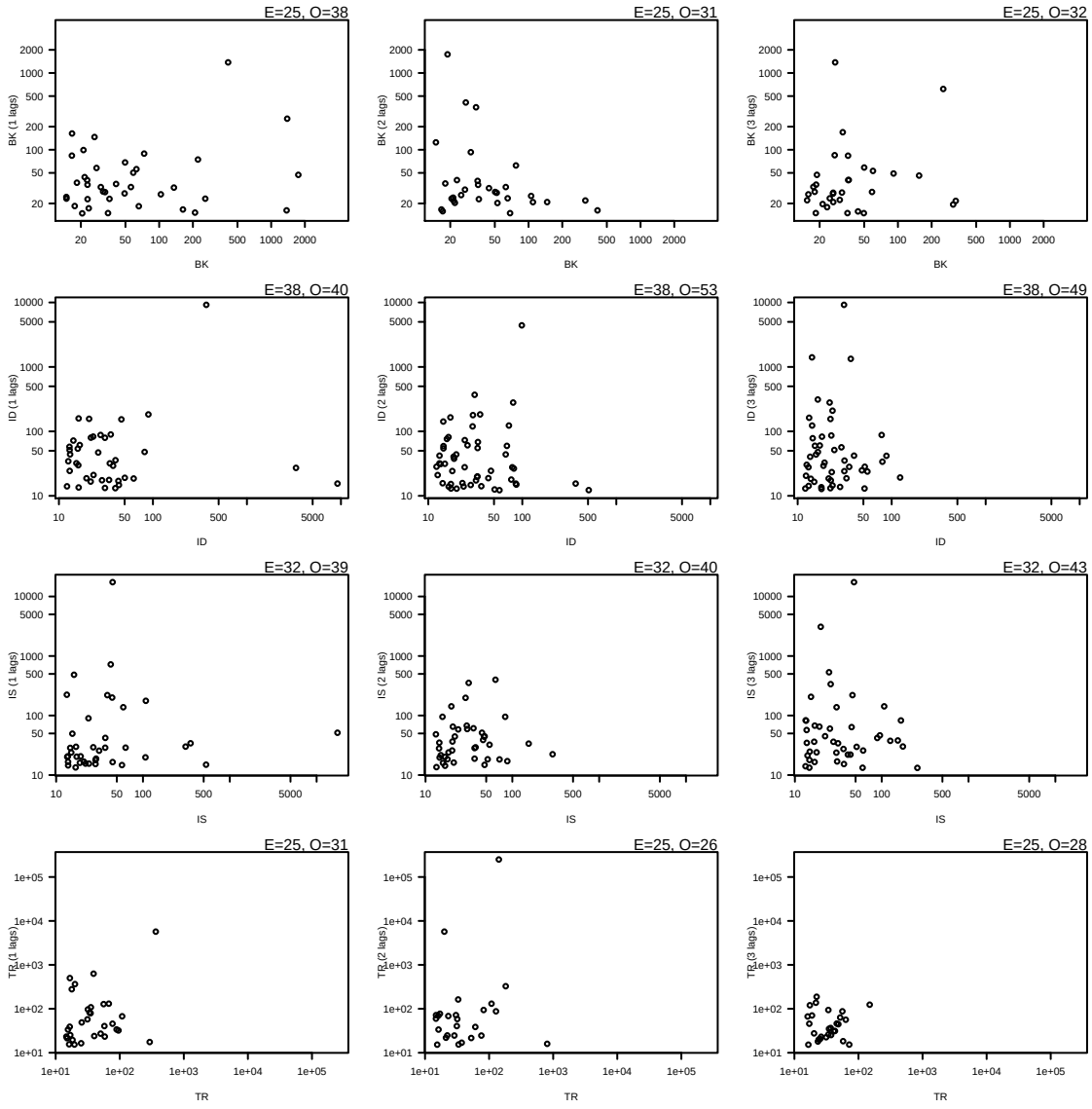

Figure 8.8: *Joint Exceedances at Different Lags of the Same Series. Variables are transformed to unit Fréchet scale. From top to bottom: BK, ID, IS, and TR. From left to right: 1, 2, and 3 lags.*

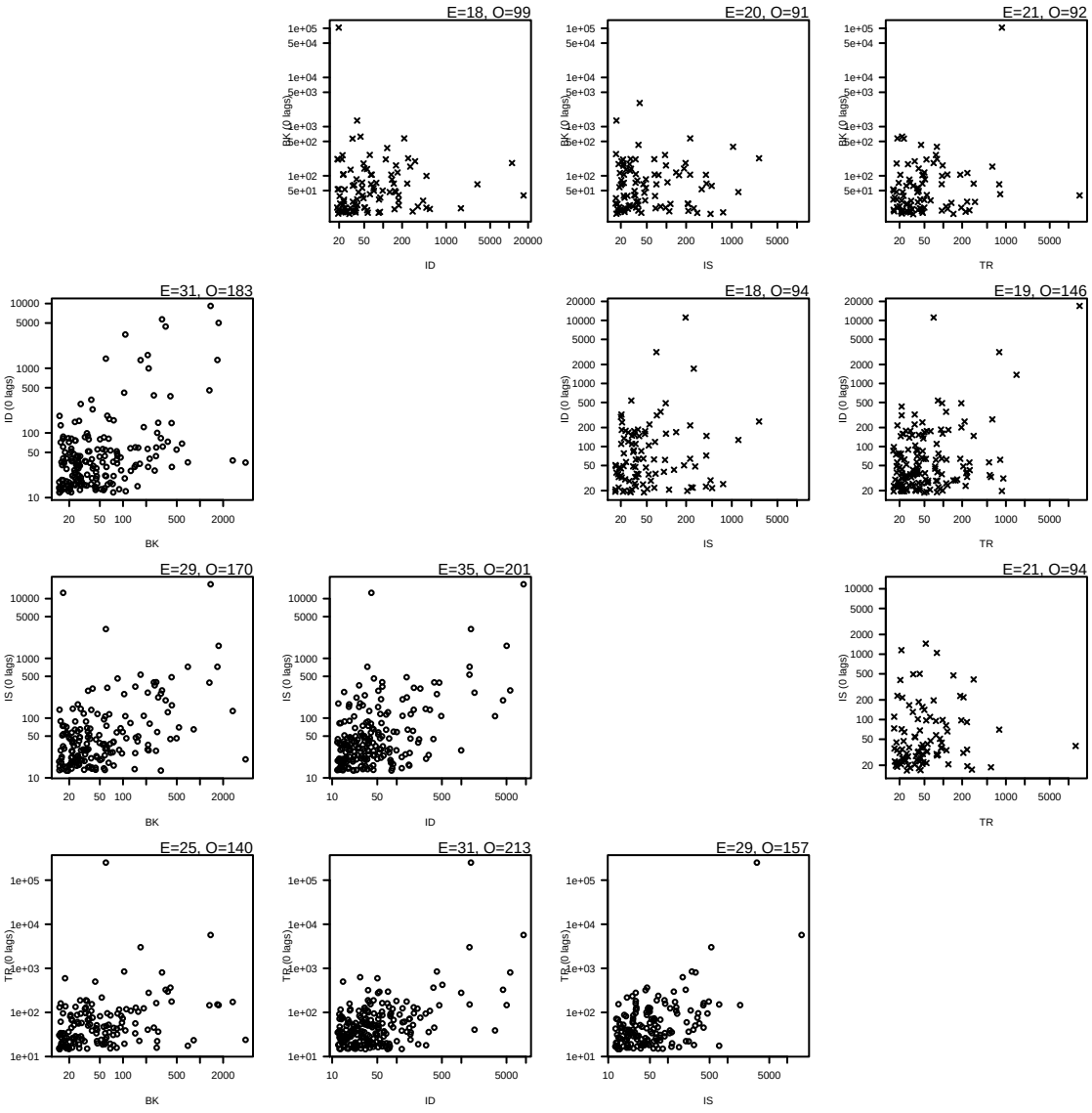

Figure 8.9: *Joint Exceedances Across Series in the Same Day. Variables are transformed to unit Fréchet scale. Lower diagonal (circles) corresponds to losses and upper diagonal (crosses) corresponds to profits.*

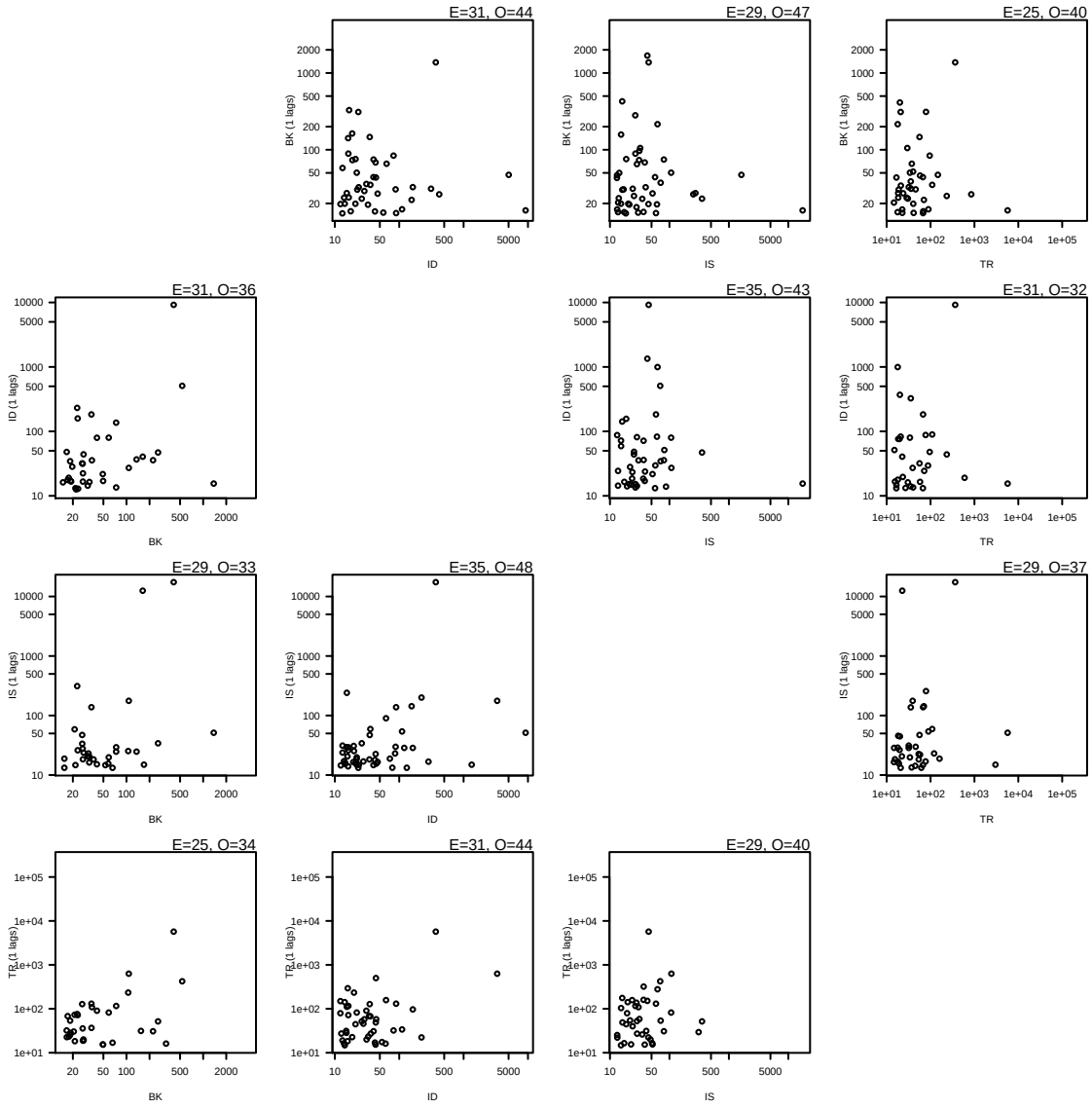

Figure 8.10: *Joint Exceedances at Different Lags Across Series. Variables are transformed to unit Fréchet scale. Lower diagonal corresponds to y_t vs. x_{t-1} (y_t in y-axis, x_{t-1} in x-axis). Upper diagonal corresponds to x_t vs. y_{t-1} (x_t in y-axis, y_{t-1} in x-axis).*

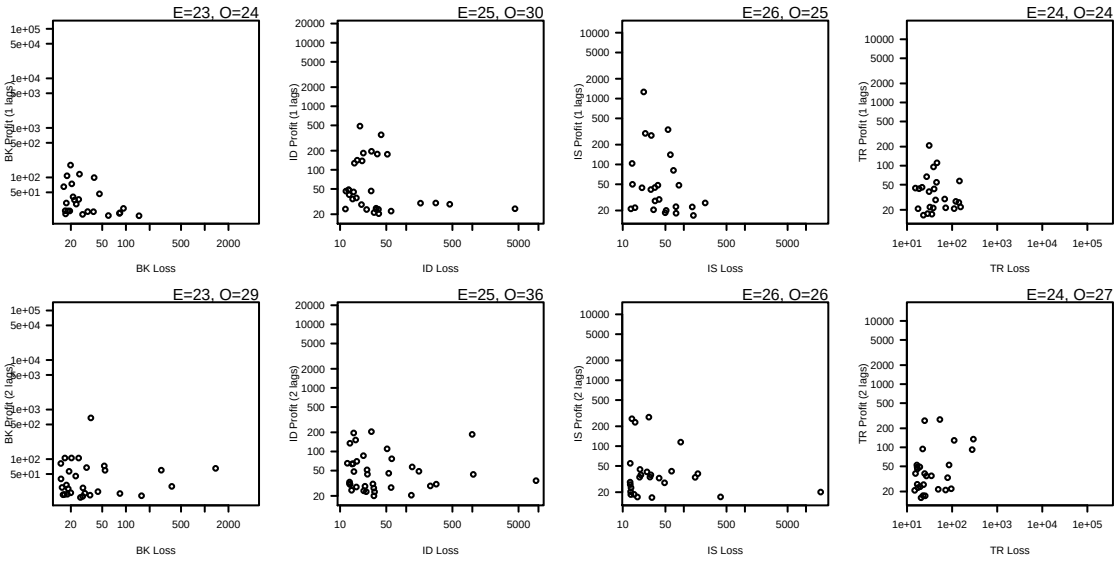

Figure 8.11: *Joint Exceedances of Upper and Lower Tails at 1 and 2 Lags. Variables are transformed to unit Fréchet scale.*

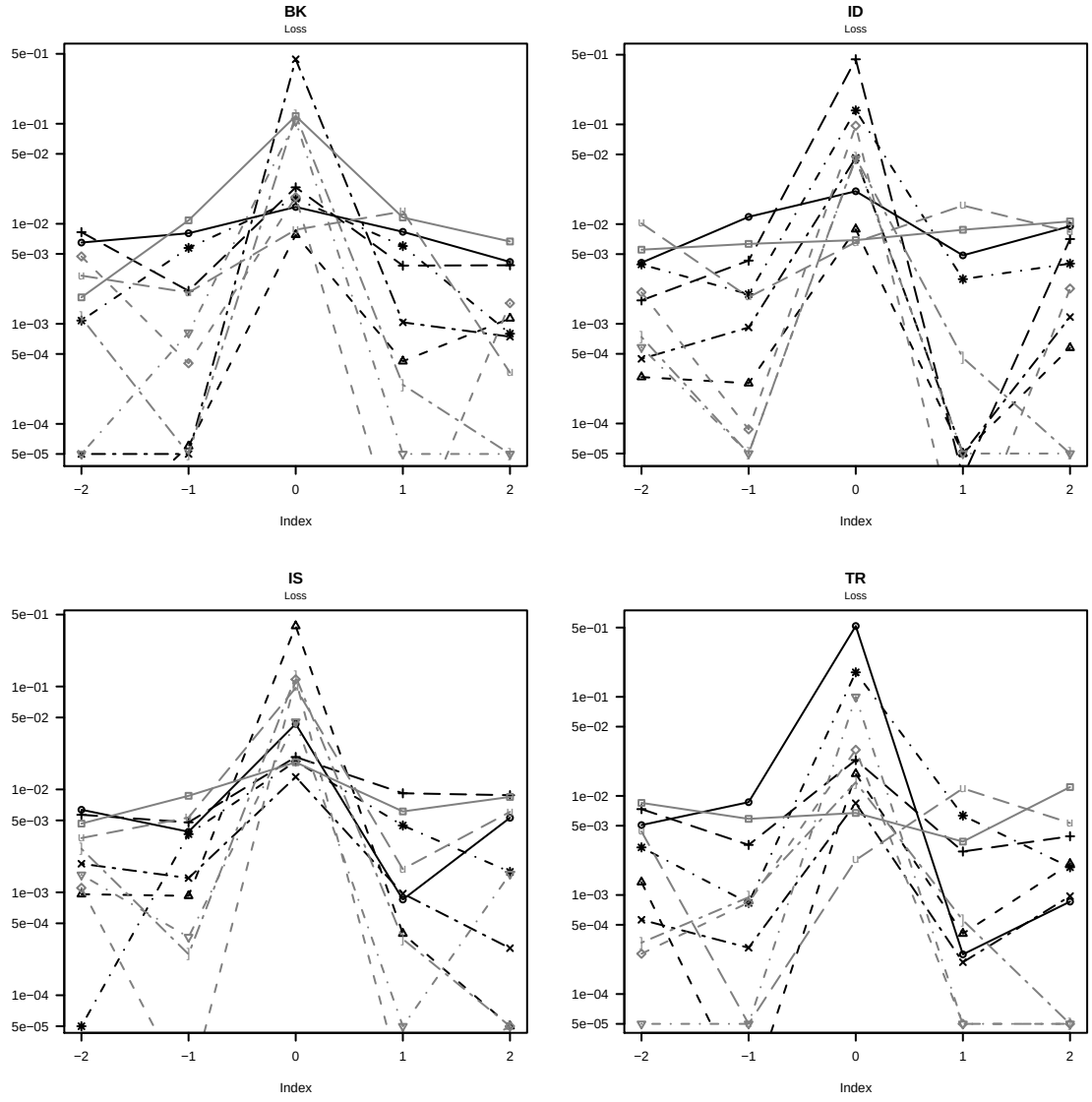

Figure 8.12: *Parameter Estimates of M4 Process Model (Losses). Model Fitted to Nasdaq Indices Standardized Residuals with Clustering-Based Method,* $-2 \leq k \leq 2$, $L = 10$.

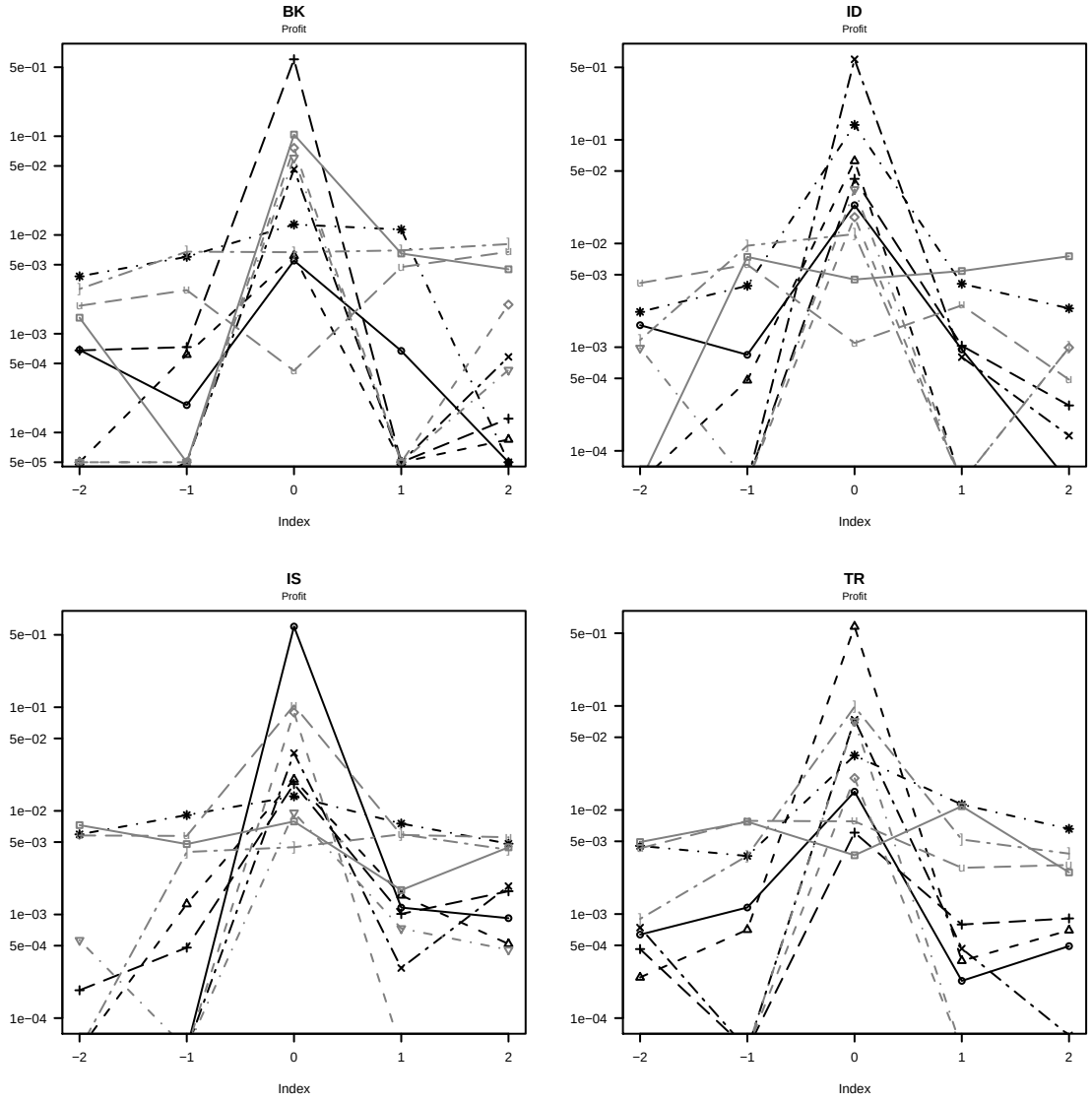

Figure 8.13: *Parameter Estimates of M4 Process Model (Profits). Model Fitted to Nasdaq Indices Standardized Residuals with Clustering-Based Method,* $-2 \leq k \leq 2$, $L = 10$.

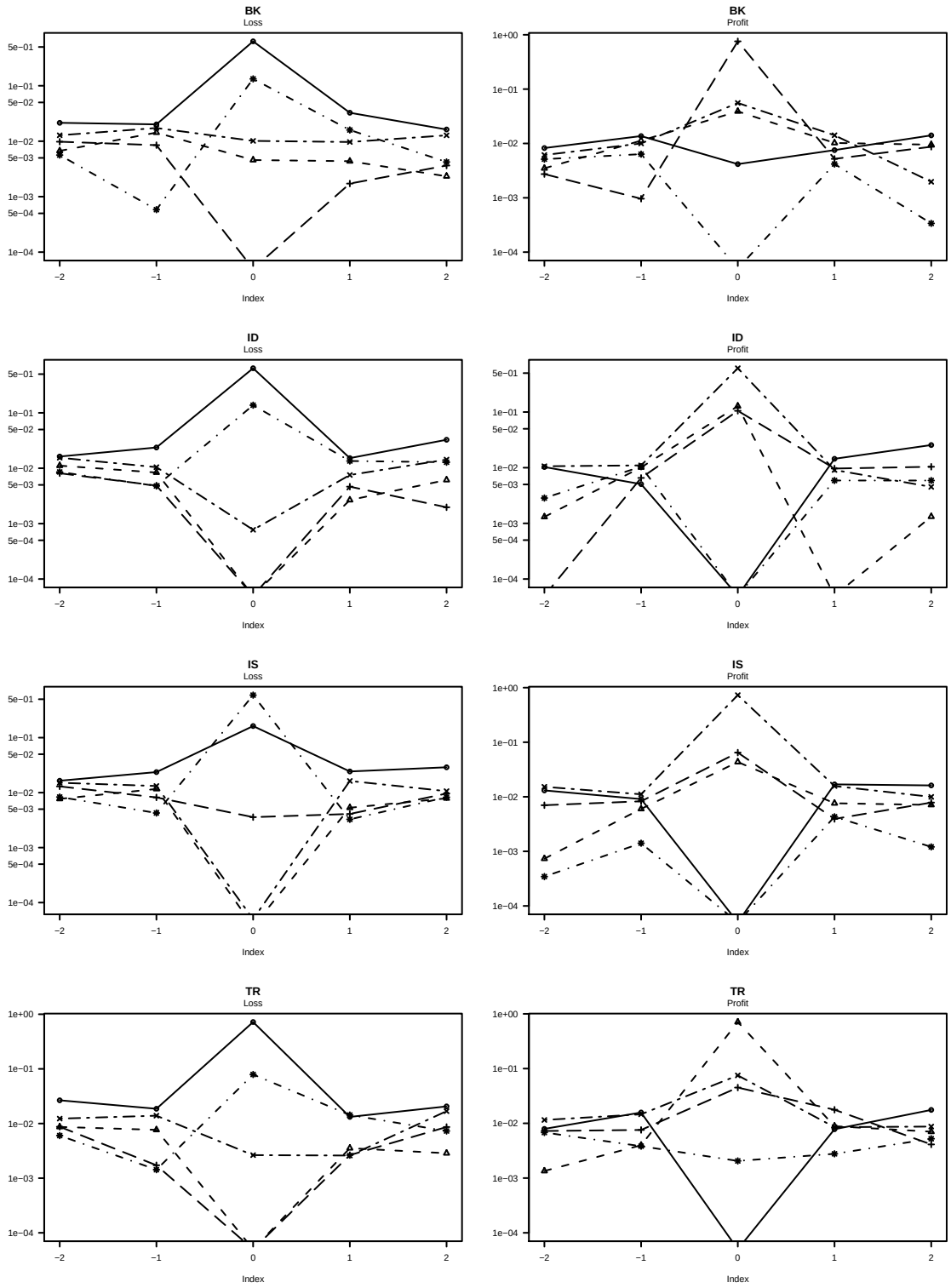

Figure 8.14: *Parameter Estimates of joint M4 Process Model for Profits and Losses. Model Fitted to Nasdaq Indices Standardized Residuals with Clustering-Based Method,* $-2 \leq k \leq 2$, $L = 5$.

APPENDIX A

Properties of the Fréchet Distribution

The Fréchet distribution is frequently used throughout this work. We state some of its properties for convenience.

The Fréchet distribution function with shape parameter $\zeta > 0$ is

$$\Phi_\zeta(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left\{-x^{-\zeta}\right\}, & x > 0. \end{cases} \tag{A.1}$$

A random variable Z with this distribution is called a Fréchet random variable. The corresponding density is

$$\varphi_\zeta(x) = \begin{cases} 0, & x \leq 0, \\ \zeta\, x^{-\zeta-1} \exp\left\{-x^{-\zeta}\right\}, & x > 0. \end{cases} \tag{A.2}$$

This density has a maximum at

$$x^* = \left(\frac{\zeta}{1+\zeta}\right)^{1/\zeta}. \tag{A.3}$$

It can be shown that all the moments of a Fréchet random variable do not exist. The quantiles x_q, $0 \leq q \leq 1$, satisfy

$$x_q = \frac{\zeta}{\log(1/q)}. \tag{A.4}$$

In particular the median is

$$x_{0.5} = \frac{\zeta}{\log 2}. \tag{A.5}$$

Unit Fréchet Distribution

A Fréchet distribution with $\zeta = 1$ is called *unit Fréchet* or *standard Fréchet*. Its distribution function is

$$\Phi_1(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left\{-\frac{1}{x}\right\}, & x > 0. \end{cases} \tag{A.6}$$

Truncated Fréchet

Let Z be a unit Fréchet random variable with scale parameter α, and let $u > 0$. The distribution function of Z given that $Z < u$, denoted $\Phi_{1,\alpha}(x \mid x < u)$, is

$$\Phi_{1,\alpha}(x \mid x < u) \equiv P\{Z < x \mid Z < u\} = \begin{cases} \exp\left\{-\frac{\alpha}{x} + \frac{\alpha}{u}\right\}, & 0 \leq x \leq u, \\ 1, & x > u. \end{cases} \tag{A.7}$$

The corresponding density function is

$$\varphi_{1,\alpha}(x \mid x < u) = \begin{cases} \frac{\alpha}{x^2} \exp\left\{-\frac{\alpha}{x} + \frac{\alpha}{u}\right\}, & 0 \leq x \leq u, \\ 0, & x > u. \end{cases} \tag{A.8}$$

Hence, the median of this distribution is

$$x_{u,0.5} = \alpha \left(\log 2 + \frac{\alpha}{u}\right)^{-1}. \tag{A.9}$$

Fréchet Exceedances

Let Z be a unit Fréchet random variable with scale parameter α, and let $u > 0$. The distribution function of Z given that $Z > u$, denoted $\Phi_{1,\alpha}(z \mid z > u)$, is

$$\Phi_{1,\alpha}(x \mid x > u) \equiv P\{Z < x \mid Z > u\} = \begin{cases} 0, & x \leq u, \\ \dfrac{\exp\left\{-\frac{\alpha}{x}\right\} - \exp\left\{-\frac{\alpha}{u}\right\}}{1 - \exp\left\{-\frac{\alpha}{u}\right\}}, & x > u. \end{cases} \tag{A.10}$$

The corresponding density function is

$$\varphi_{1,\alpha}(x \mid x > u) = \begin{cases} 0, & x \leq u, \\ \dfrac{\alpha}{x^2} \dfrac{\exp\left\{-\frac{\alpha}{x}\right\}}{1 - \exp\left\{-\frac{\alpha}{u}\right\}}, & x > u. \end{cases} \tag{A.11}$$

For $x > u$ the median of this distribution, denoted $m(u)$, satisfies

$$\frac{\exp\left\{-\frac{\alpha}{m(u)}\right\} - \exp\left\{-\frac{\alpha}{u}\right\}}{1 - \exp\left\{-\frac{\alpha}{u}\right\}} = \frac{1}{2}, \tag{A.12}$$

from which we obtain

$$m(u) = \frac{\alpha}{\log 2 - \log\left(1 + \exp\left\{-\frac{\alpha}{u}\right\}\right)}. \tag{A.13}$$

A first-order Taylor expansion of (A.13) as $u \to \infty$ gives

$$m(u) = 2u + \frac{\alpha}{2} + \frac{\alpha^2}{8u} + O\left(\frac{1}{u^2}\right). \tag{A.14}$$

For sufficiently large u, a rough approximation of $m(u)$ is given by

$$m(u) \approx 2u + \frac{\alpha}{2}. \tag{A.15}$$

Simulation

Algorithm A.0.1. The following generates a random variable Z with distribution function (A.7).

1. Generate $U \sim \mathcal{U}(0, e^{-\alpha/u})$.
2. Set $Z = -\alpha / \log U$.

Proof.

$$\begin{aligned} P\{Z \leq z\} &= P\{\log U \leq -\alpha/z\} \\ &= P\left\{U \leq e^{-\alpha/z}\right\} \end{aligned}$$

$$= \Phi_{1,\alpha}\left(z \mid z < u\right). \tag{A.16}$$

□

Algorithm A.0.2. The following generates a random variable Z with distribution function (A.10).

1. Generate $U \sim \mathcal{U}(e^{-\alpha/u}, 1)$.
2. Set $Z = -\alpha/\log U$.

Proof.

$$\begin{aligned} P\left\{Z \leq z\right\} &= P\left\{\log U \leq -\alpha/z\right\} \\ &= P\left\{U \leq e^{-\alpha/z}\right\} \\ &= \Phi_{1,\alpha}\left(z \mid z > u\right). \end{aligned} \tag{A.17}$$

□

APPENDIX B

Sampling Techniques

Consider a function $f(\cdot) > 0$, such that $\int f(x)\mathrm{d}x < \infty$. Suppose we can sample from a density $g(\cdot)$ and we know $f(\cdot)$ up to proportionality. The task is to simulate samples from

$$h(x) = \frac{f(x)}{\int f(x')\mathrm{d}x'}. \tag{B.1}$$

We review two different approaches to solve this problem: importance sampling, and weighted bootstrap (sampling/importance resampling).

B.1 Importance Sampling

The goal is to estimate

$$\mathbb{E}\left(\phi(x)\right) = \int \phi(x)h(x)\mathrm{d}x. \tag{B.2}$$

Define $w(x) = f(x)/g(x)$, then we can write

$$\begin{aligned}\mathbb{E}\left(\phi(x)\right) &= \frac{\int \phi(x)f(x)\mathrm{d}x}{\int f(x)\mathrm{d}x} \\ &= \frac{\int \phi(x)w(x)g(x)\mathrm{d}x}{\int w(x)g(x)\mathrm{d}x}\end{aligned} \tag{B.3}$$

If we obtain a sample $X_1, X_2, \ldots, X_n \sim g(\cdot)$ then we can estimate

$$\begin{aligned}\widehat{\mathbb{E}\left(\phi(x)\right)} &= \frac{\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)w(x_i)}{\frac{1}{N}\sum_{i=1}^{N}w(x_i)} \\ &= \frac{\sum_{i=1}^{N}\phi(x_i)w(x_i)}{\sum_{i=1}^{N}w(x_i)}.\end{aligned} \tag{B.4}$$

In other words,

$$F(x) = P\{X \leq x\} = \int \mathbb{I}(X \leq x)\, h(x)\mathrm{d}x \approx \frac{\sum_{i=1}^{N} \mathbb{I}(X \leq x)\, w(x_i)}{\sum_{i=1}^{N} w(x_i)}, \tag{B.5}$$

so that

$$\widehat{h(x)} \approx \frac{\sum_{i=1}^{N} w(x_i)\delta(x_i - x)}{\sum_{i=1}^{N} w(x_i)}, \tag{B.6}$$

where $\delta(\cdot)$ is the Dirac delta measure

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ 1, & x = 0. \end{cases} \tag{B.7}$$

The following example illustrates importance sampling, and it also motivates an improved particle filtering algorithm for MM processes.

Example B.1.1 (Importance Sampling for Tail Probabilities). Let Z be a unit Fréchet random variable with density $\varphi_1(x)$. The goal is to estimate, for large z^*,

$$\zeta = P\{Z > z^*\} = \int_0^\infty \mathbb{I}(x > z^*)\, \varphi_1(x)\mathrm{d}x. \tag{B.8}$$

The true value is $\zeta = 1 - e^{-1/z^*}$. For example, when $z^* = 100$, $\zeta = 0.00995$.

We compare two approaches to this problem, one based on direct sampling, and a second based on importance sampling.

Direct Approach: Sample $z_i \sim \Phi_1(x),\ i = 1, \ldots, N$, then

$$\hat{\zeta} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(z_i > z^*). \tag{B.9}$$

However, this estimate converges very slowly because any sample will have very few observations in the tail.

Importance Sampling Approach: Denote by $\varphi_1(x \mid x > u)$ the density of the excess

distribution above a high threshold u of a unit Fréchet random variable as defined in (A.10). Define

$$w(x) = \frac{\varphi_1(x)}{\varphi_1(x \mid x > u)} = \begin{cases} \infty, & x \leq u, \\ 1 - e^{-1/u}, & x > u. \end{cases} \tag{B.10}$$

For $z^* > u$ we can write

$$\begin{aligned} \zeta &= \int_0^\infty \mathbb{I}\left(x > z^*\right) \varphi_1(x) \mathrm{d}x \\ &= \int_0^\infty \mathbb{I}\left(x > z^*\right) w(x) \varphi_1(x \mid x > u) \mathrm{d}x \end{aligned} \tag{B.11}$$

Therefore, we can obtain a sample $z_i \sim \Phi_1\left(x \mid x > u\right),\ i = 1, \ldots, N$, and then

$$\begin{aligned} \hat{\zeta} &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(z_i > z^*\right) w(z_i) \\ &= \left(1 - e^{-1/u}\right) \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(z_i > z^*\right). \end{aligned} \tag{B.12}$$

The weights do not have to be standardized because the normalizing constant for $\varphi_1(x)$ is known. If $u = z^*$ then $\mathbb{I}\left(z_i > z^*\right) = 1$ for all i, so the estimate is the true value regardless of the sample size.

Figure B.1 illustrates the estimates obtained with both methods as a function of the sample size. The importance sampling estimate converges to the true value around $N = 100$. The direct method converges very slowly to the true value and the variance of the estimate is much larger than the variance of the importance sampling method. ‖

B.2 Weighted Bootstrap

The weighted bootstrap algorithm is described in Smith and Gelfand (1992). This is a variant of the well known bootstrap (Efron, 1982). Rubin (1988) refers to this procedure as sampling/importance resampling (SIR). The algorithm is as follows.

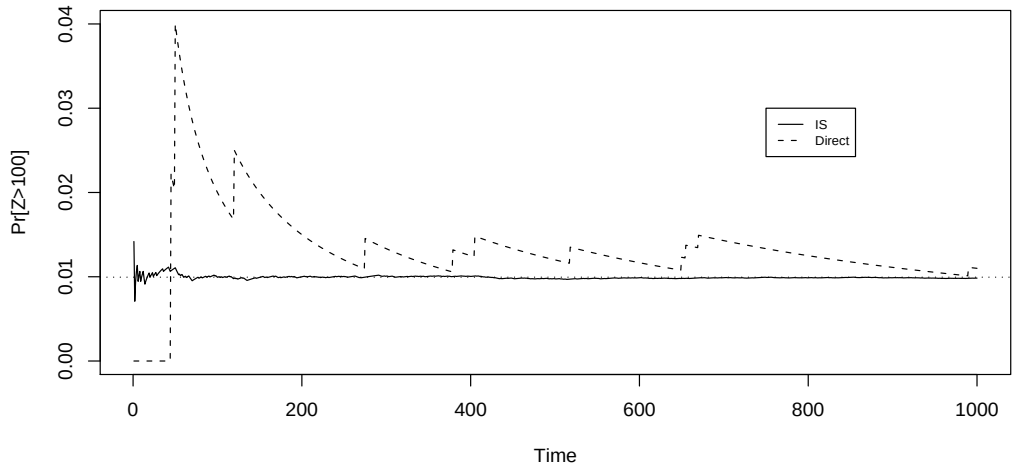

Figure B.1: *Estimates of Tail Probability.*

1. Simulate a sample from $g(\cdot)$

$$\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(N)} \sim g(\cdot).$$

2. For each $\tilde{x}^{(i)},\ i = 1, \ldots, N$ calculate

$$\tilde{w}^{(i)} = \frac{f(\tilde{x}^{(i)})}{g(\tilde{x}^{(i)})}, \qquad w^{(i)} = \frac{\tilde{w}^{(i)}}{\sum_{i=1}^{N} \tilde{w}^{(i)}}.$$

3. Draw samples with replacement from the set $\left\{\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(N)}\right\}$, according to the probabilities $w^{(i)}$ to obtain a sample

$$x^{(1)}, x^{(2)}, \ldots, x^{(N)}.$$

   This sample is approximately distributed as $h(x)$.

A modification described by Rubin (1988) is to start with $R > N$ samples and then resample N of them. Rubin (1988) explains that the rationale for this is that as $R/N \to \infty$, the N values $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$ are drawn with probabilities given by

$$g(x)\frac{w^{(i)}(x)}{\int g(x)w(x)\mathrm{d}x} = \frac{f(x)}{\int_{-\infty}^{\infty} f(x)\mathrm{d}x} = h(x).$$

Justification of the Weighted Bootstrap Algorithm

Suppose the random variable X is obtained with the algorithm described above. Given the sample $\left\{\tilde{x}^{(i)}\right\},\ i = 1, \ldots, N$ from $g(\cdot)$, we have

$$P\left\{ X = \tilde{x}^{(i)} \,\middle|\, \tilde{x}^{(i)},\, w^{(i)} \right\} = w^{(i)}.$$

Following the argument in Smith and Gelfand (1992),

$$\begin{aligned} P\left\{ X \leq x \,|\, \tilde{x}^{(i)},\, w^{(i)} \right\} &= \mathbb{E}\left(\mathbb{I}\left(X \leq x\right) \,\middle|\, \tilde{x}^{(i)},\, w^{(i)} \right) \\ &= \sum_{i=1}^{N} w^{(i)} \mathbb{I}\left(\tilde{x}^{(i)} \leq x \right) \\ &= \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}^{(i)} \mathbb{I}\left(\tilde{x}^{(i)} \leq x\right)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}^{(i)}}. \end{aligned} \tag{B.13}$$

The numerator in (B.13) is the sample mean of the iid random variables $\tilde{w}^{(i)} \mathbb{I}\left(\tilde{x}^{(i)} \leq x\right)$, hence by the strong law of large numbers the numerator converges to

$$\begin{aligned} \mathbb{E}\left(\tilde{w}^{(i)} \mathbb{I}\left(\tilde{x}^{(i)} \leq x \right) \right) &= \int \frac{f(u)}{g(u)} \mathbb{I}\left(u \leq x\right) g(u) \mathrm{d}u \\ &= \int_{-\infty}^{x} f(u) \mathrm{d}u. \end{aligned}$$

Similarly, the denominator converges to $\int_{-\infty}^{\infty} f(u)\mathrm{d}u$. Therefore, as $N \to \infty$

$$\begin{aligned} \mathbb{E}\left(\mathbb{I}\left(X \leq x\right) \,\middle|\, \tilde{x}^{(i)},\ w^{(i)} \right) &\overset{\text{a.s.}}{\to} \frac{\int \frac{f(u)}{g(u)} \mathbb{I}\left(u \leq x\right) g(u) \mathrm{d}u}{\int \frac{f(u)}{g(u)} g(u) \mathrm{d}u} \\ &= \frac{\int_{-\infty}^{x} f(u)\mathrm{d}u}{\int_{-\infty}^{\infty} f(u)\mathrm{d}u} \\ &= \int_{-\infty}^{x} h(u)\mathrm{d}u. \end{aligned}$$

Therefore, conditionally on the sample from $g(\cdot)$, as $N \to \infty$, X will be approximately distributed as $h(\cdot)$.

B.3 Dirichlet Distribution

Following Gelman *et al.* (1995, p. 76), a random vector $\boldsymbol{X} = (X_1, \ldots, X_k)'$, satisfying $X_j \geq 0$ and $\sum_j X_j = 1$ is said to have a Dirichlet distribution with parameters $\theta_1, \ldots, \theta_k$, where $\theta_j > 0$ if $\boldsymbol{X}$ has the density function

$$f(\boldsymbol{x}) = \frac{\Gamma(\theta_1 + \cdots + \theta_k)}{\Gamma(\theta_1) \cdots \Gamma(\theta_k)} x_1^{\theta_1} \cdots x_k^{\theta_k}, \qquad x_j \geq 0, \ \sum_j x_j = 1. \tag{B.14}$$

Define $\theta_0 = \sum_{j=1}^k \theta_j$. The first- and second-order moments of $\boldsymbol{X}$, and its mode are

$$\begin{aligned}
\mathbb{E}(X_j) &= \frac{\theta_j}{\theta_0} && \text{(B.15)} \\
\text{Var}(X_j) &= \frac{\theta_j(\theta_0 - \theta_j)}{\theta_0^2(\theta_0 + 1)} && \text{(B.16)} \\
\text{Cov}(X_i, X_j) &= -\frac{\theta_i \theta_j}{\theta_0^2(\theta_0 + 1)} && \text{(B.17)} \\
\text{mode}(X_j) &= \frac{\theta_j - 1}{\theta_0 - k} && \text{(B.18)}
\end{aligned}$$

If $\theta_j = 1$ for all j then $\boldsymbol{X}$ has a uniform distribution; i.e., a distribution that asigns the same density to any vector $\boldsymbol{x}$ satisfying $x_j \geq 0$ and $\sum_j x_j = 1$. In this case the density is

$$f(\boldsymbol{x}) = (k-1)!, \qquad x_j \geq 0, \ \sum_j x_j = 1. \tag{B.19}$$

The mode is not defined (is of the form 0/0), and the first- and second-order moments are

$$\begin{aligned}
\mathbb{E}(X_j) &= \frac{1}{k} && \text{(B.20)} \\
\text{Var}(X_j) &= \frac{k-1}{k^2(k+1)} && \text{(B.21)} \\
\text{Cov}(X_i, X_j) &= -\frac{1}{k^2(k+1)} && \text{(B.22)}
\end{aligned}$$

BIBLIOGRAPHY

Ait-Sahalia Y. and Hansen L.P. (editors) (2005). *Handbook of Financial Econometrics*. North Holland. Forthcoming.

Andersen T.G. and Sorensen B.E. (1996). GMM estimation of a stochastic volatility model: a Monte Carlo study. *J. Bus. Econom. Statist.* **14**, 28–352.

Artzner P., Delbaen F., Eber J. and Heath D. (1999). Coherent measures of risk. *Math. Finance* **9**, 203–228.

Arulampalam M.S., Maskell S., Gordon N. and Clapp T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50**, 174–188.

Beirlant J., Goegebeur Y., Segers J. and Teugels J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

BIS (1993). *Payment Systems in the Group of Ten Countries*. Bank for International Settlements, Basle. Prepared by the Committee on Payment and Settlement Systems of the central banks of the Group of Ten countries.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

Bollerslev T., Engle R.F. and Nelson D.B. (1994). ARCH models. In: *Handbook of Econometrics, Volume IV*, edited by R.F. Engle and D.L. McFadden, pp. 2959–3038. Elsevier Science B.V.

Brockwell P.J. and Davis R.A. (1993). *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition.

Campbell J.Y., Lo A.W. and MacKinlay A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey.

Carter C.K. and Kohn R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–553.

Coles S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Coles S.G. (2003). The use and misuse of extreme value models in practice. In: Finkenstädt and Rootzén (2003), pp. 79–100.

Crisan D. and Doucet A. (2002). A survey of convergence results on particle filtering for practitioners. *IEEE Trans. Signal Process.* **50**, 736–746.

Daníelsson J. and de Vries C. (2001). Value at risk and extreme returns. FMG-Discussion Paper 273, Financial Markets Group, London School of Economics.

Daníelsson J., Embrechts P., Goodhart C., Keating C., Muennich F., Renault O. and Shin H.S. (2001). An academic response to Basel II. FMG Special Papers sp0130, Financial Markets Group and ESRC.

Davis R.A. and Resnick S.I. (1989). Basic properties and prediction of max-ARMA processes. *Adv. Appl. Probab.* **21**, 781–803.

Deheuvels P. (1983). Point processes and multivariate extreme values. *J. Multivariate Anal.* **13**, 257–272.

Doucet A., De Freitas N. and Gordon N. (editors) (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Doucet A. and Tadić V.B. (2003). Parameter estimation in general state-space models using particle methods. *Ann. Inst. Statist. Math.* **55**, 409–422.

Efron B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.

Embrechts P., Klüppelberg C. and Mikosch T. (1997). *Modelling Extremal Events: for Insurance and Finance*. Springer, New York.

Engle R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

Engle R.F. (2001). Financial econometrics – a new discipline with new methods. *J. Econometrics* **100**, 53–56.

Finkenstädt B. and Rootzén H. (editors) (2003). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall/CRC Press, Boca Raton.

Fougères A. (2003). Multivariate extremes. In: Finkenstädt and Rootzén (2003), pp. 373–388.

Galambos J. (1987). *Asymptotic Theory of Extreme Order Statistics*. R.E. Krieger Pub. Co., Malabar, 2nd edition.

Garcia R., Ghysels E. and Renault E. (2005). The econometrics of option pricing. In: Ait-Sahalia and Hansen (2005). Forthcoming.

Gelfand A.E. and Smith A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Gelman A., Carlin J.B., Stern H.S. and Rubin D.B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Geman S. and Geman D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741.

Ghysels E., Santa-Clara P. and Valkanov R. (2004a). Predicting volatility: Getting the most out of return data sampled at different frequencies. J. Econometrics (forthcoming).

Ghysels E., Santa-Clara P. and Valkanov R. (2004b). There is a risk-return tradeoff after all. J. Financial Economics (forthcoming).

Gilks W.R., Richardson S. and Spiegelhalter D.J. (editors) (1996). *Markov Chain Monte Carlo Methods in Practice*. Chapman & Hall, London.

Gordon N.J., Salmond D.J. and Smith A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Proc. IEE-F* **140**, 107–113. Radar and Signal Processing.

Hall P., Peng L. and Yao Q. (2002). Moving-maximum models for extremes of time series. *J. Statist. Plann. Inference* **103**, 51–63.

Hartigan J. (1975). *Clustering Algorithms*. Wiley, New York.

Harvey A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey A.C., Ruiz E. and Shephard N. (1994). Multivariate stochastic variance models. *Rev. Econom. Stud.* **61**, 247–264.

Hastings W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika* .

Hürzeler M. and Künsch H.R. (2001). Approximating and maximising the likelihood for a general state-space model. In: Doucet *et al.* (2001), pp. 159–175.

Jacquier E., Polson N.G. and Rossi P.E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econom. Statist.* **12**, 371–417.

Kalman R.E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng. Ser. D* **82**, 35–45.

Kitagawa G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5**, 1–25.

Kotz S. and Nadarajah S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.

Leadbetter M.R., Lindgren G. and Rootzén H. (1983). *Extremes and Related Properties of Random Sequences and Series*. Springer-Verlag, New York.

Liu J. and West M. (2001). Combined parameter and state estimation in simulation-based filtering. In: Doucet *et al.* (2001), pp. 197–223.

McNeil A.J. and Frey R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empirical Finance* **7**, 271–300.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

Nandagopalan S. (1990). *Multivariate Extremes and the Estimation of the Extremal Index*. PhD dissertation, Department of Statistics, University of North Carolina, Chapel Hill, NC.

Nandagopalan S. (1994). On the multivariate extremal index. *J. Res. Nat. Inst. Standards Tech.* **99**, 543–550.

Pitt M.K. and Shephard N. (1999a). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**, 590–599.

Pitt M.K. and Shephard N. (1999b). Time-varying covariances: a factor stochastic volatility approach. In: *Bayesian Statistics 6*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 547–570. Oxford University Press, Oxford.

Pitt M.K. and Shephard N. (2001). Auxiliary variable based particle filters. In: Doucet *et al.* (2001), pp. 273–293.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Resnick S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.

Robert C.P. and Casella G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition.

Rubin D.B. (1988). Using the SIR algorithm to simulate posterior distributions. In: *Bayesian Statistics 3*, edited by J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, pp. 395–402. Oxford University Press, Oxford.

Shephard N. (1996). Statistical aspects of ARCH and stochastic volatility. In: *Time Series Models: in Econometrics, Finance and Other Fields*, edited by D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, pp. 1–67. Chapman & Hall, London.

Shephard N. and Pitt M.K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–667.

Smith A.F.M. and Gelfand A.E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *Amer. Statist.* **46**, 84–88.

Smith R.L. (2003). Statistics of extremes, with applications in environment, insurance, and finance. In: Finkenstädt and Rootzén (2003), pp. 1–78.

Smith R.L., Tawn J.A. and Coles S.G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268.

Smith R.L., Tawn J.A. and Yuen H.K. (1990). Statistics of multivariate extremes. *Internat. Statist. Rev.* **58**, 47–58.

Smith R.L. and Weissman I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56**, 515–528.

Smith R.L. and Weissman I. (1996). Characterization and estimation of the multivariate extremal index. Preprint, under revision.

Taylor S.J. (1986). *Modelling Financial Time Series*. John Wiley, Chichester.

West M. and Harrison P.J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd edition.

Zhang Z. (2002). *Multivariate Extremes, Max-Stable Process Estimation and Dynamic Financial Modeling*. PhD dissertation, Department of Statistics, University of North Carolina, Chapel Hill, NC.

Zhang Z. and Smith R.L. (2001). Modeling financial time series data as moving maxima processes. Manuscript, UNC.

Zhang Z. and Smith R.L. (2003). The behavior of multivariate maxima of moving maxima processes. Submitted for publication.